RNA STRUCTURE PROBING

biochemical structure analysis of autoimmune-related RNA molecules

een wetenschappelijke proeve op het gebied van de
Natuurwetenschappen, Wiskunde en Informatica

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
maandag 6 december 1999,
des namiddags om 1.30 uur precies
door

ALEXANDER WILHELMUS MARIA TEUNISSEN

geboren op 21 september 1971
te Nijmegen

Promotor: Prof. Dr. W. J. van Venrooij

Co-promotor: Dr. G. J. M. Pruijn

Manuscriptcommissie: Prof. Dr. C. W. Hilbers
Prof. Dr. C. W. Pleij (RU Leiden)

ISBN: 90-9013234-1

Contents

1

GENERAL INTRODUCTION

List of abbreviations

3'UTR	3'untranslated region
CMCT	1-cyclohexyl-3(2-morpholinoethyl) carbodiimide methane-p-toluene sulfonate
DEPC	diethylpyrocarbonate
DMS	dimethyl sulphate
ds	double-stranded
dsRBD	double-stranded RNA binding motif
ENU	ethylnitrosourea
FRET	Fluorescence Resonance Energy Transfer
hnRNP	heterogeneous nuclear RNP
KH domain	K homology domain
NMR	nuclear magnetic resonance
RNase	ribonuclease
RNP	ribonucleoprotein particle
RNP motif	RNA binding motif
RRE	rev-responsive element
RRM	RNA recognition motif
snRNP	small nuclear RNP
scFv	single chain variable fragment
SLE	Systemic lupus erythematosus
ss	single-stranded

Introduction to RNA structure

RNA structure and RNA-protein interactions

Our knowledge on the function and role of RNA in cellular processes has increased dramatically in the past decade. Apart from being the molecule transferring genetic information from DNA to proteins, RNA molecules have been found to serve several additional functions. Although RNA with its four basic building blocks seems less complex than for example proteins, which have twenty basic building blocks, it is surprisingly versatile in structure formation. The primary structure (one-dimensional: sequence) determines the secondary structure (two-dimensional), which folds into the tertiary (three-dimensional) structure of the RNA. The three-dimensional folding of the RNA will ultimately determine the function of the RNA.

RNA secondary structure elements - RNA molecules are synthesised as single-stranded molecules. Most structural RNAs show high levels of base pairing within the molecule which probably stabilises the RNA to prevent premature degradation. The base pairing patterns can be classified in different secondary structure elements. The stem, bulge, internal loop, hairpin, and junction are common RNA secondary structure motifs [1] (shown in Figure 1). They stabilise the RNA structure and function as basic building blocks around which the tertiary folding can be achieved. Apart from the Watson-Crick or canonical base pairing, several other types of interaction can be found. Stacking of helices, Hoogsteen base pairing and other tertiary interactions ultimately contribute to the three-dimensional folding and hence to the function of the RNA.[2,3] The diversity of different three-dimensional folds increases with each experimentally determined structure and some of the more important ones will be discussed in more detail. Since the first X-ray structure of tRNA,[4,5] almost every newly determined structure showed a novel type of tertiary interaction playing a key role in structure formation and hence in function.

Non-canonical base pairs - The refined X-ray structure of tRNA revealed for the first time the phenomenon of a non-canonical G-U base pair.[5] Around 1990, four of the five duplex RNA structures solved by crystallisation also revealed non-canonical base pairs, suggesting that RNA is able to accommodate mismatches more easily than DNA.[6-9] Not only can the RNA duplex be quite easily deformed,[10] it also displays great flexibility by adapting completely different structures in crystals compared to the solution structure as determined by NMR.[6] With the aid of more advanced techniques to obtain crystals,[11] additional three-dimensional crystallographic structures could be determined. The hammerhead ribozyme structure [12,13] and the P4-P6 domain of the group I intron[3] provided invaluable information on RNA folding. New motifs which enabled RNA to fold into very complex and, more importantly, uniquely identifiable structures were discovered in which non-canonical base pairs play an important role.[14,15]

The major groove of helical A-form RNA (the most usual conformation adopted by double-stranded RNA) is deep and narrow. Contacts of nucleotides in the major

groove with other nucleic acids and/or proteins are difficult, although the diversity of hydrogen bond donors and acceptors is greater in the major groove than in the minor groove. The latter is more accessible but mostly harbours less specific chemical groups that can be used in unique recognition of the RNA. This is where the non-canonical base pairs come into play. Non-canonical base pairs can distort the regular helical shape and hence promote widening of the major groove, thus facilitating the presentation of chemical groups in either the major or the minor groove for specific recognition or conformation.[2,16,17] This was recently exemplified in the case of selenoprotein translation, which depends on the presence of a specific structure in the 3'UTR.[18] Two sheared G-A base pairs form a structural motif necessary for selenoprotein translation.[19]

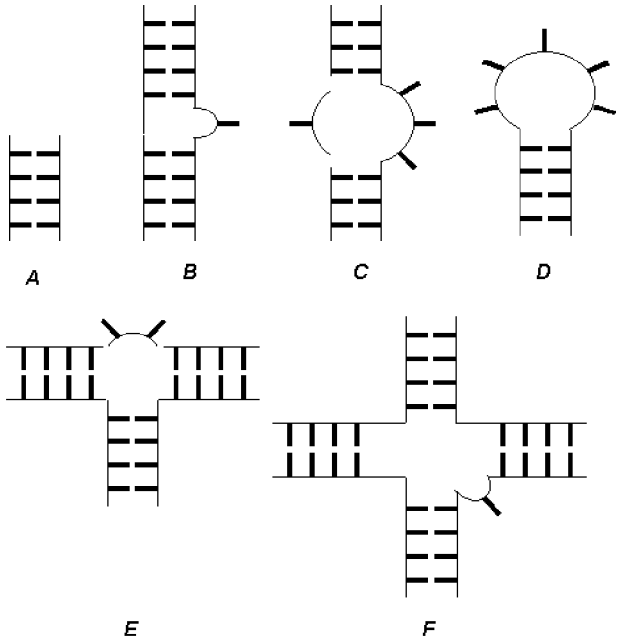


***Figure 1. Common RNA secondary structure elements**: (A) Stem structure (also called helix, duplex or double-stranded region). (B) Bulge. (C) Internal loop. Shown is an asymmetric internal loop. (D) Stem loop or hairpin. (E) and (F) show a three- and four way junction, respectively.*

2'-hydroxyl group and RNA structure - Apart from non-canonical base pairs, RNA also uses the 2'-hydroxyl group to its advantage. The 2'-hydroxyl group can act either as a hydrogen bond donor or as a hydrogen bond acceptor in the minor groove and the ribose ring can present the hydroxyl group via different conformations.[20] The *Tetrahymena* group I intron shows that this small chemical group is important for docking of the P1 helix into the active site of the group I intron.[2] The 2'-hydroxyl group also plays a major role in the formation of a *ribose zipper*, which stabilises the close packing of helices as shown in the P4-P6 domain X-ray structure.

Metal ions and RNA structure - Since the determination of the structure of tRNA [4,5] it is known that metal ions are essential for three-dimensional folding through co-ordination with negatively charged phosphate groups. Subsequent crystallographic studies have supported and extended these observations.[21-33] The importance of positively charged metal ions as scaffolds for tertiary structure formation was shown even more clearly in studies of RNA structure in solution. Ionic conditions in solution are easily manipulated, hence the dynamics of RNA structure relative to the concentration of, for example Mg^{2+} ions, can be readily monitored. In the folding of a four way RNA junction, Mg^{2+} can induce structure rearrangements from a 90° angle (low magnesium) to an anti-parallel conformation (high magnesium) of the four arms.[25] A recent study of the Tinoco group provides another example of magnesium dependent structural rearrangements.[21] In this study not only tertiary, but also the secondary folding was found to be altered by divalent cations. It was also shown that, in contrast to the common assumption, first the tertiary structure can be formed which subsequently can induce a secondary structure rearrangement.

Tertiary folding motifs - Apart from the major RNA structure building elements described above (secondary structure elements, non-canonical base pairs, 2'-hydroxyl, and metal ions), combinations of several additional folding motifs can be used to obtain a unique tertiary structure. Only a few of the more important motifs will be discussed here. A structural motif believed to be widely used in RNA, is the *tetraloop receptor motif*.[2,14,15] A stable tetraloop of the GNRA type (N: any nucleotide, R: pyrimidine) can bind to a conserved secondary structure of eleven nucleotides, forming a tertiary interaction. This motif has been found in the structures of group I and II introns [34-36] and in RNase P.[37] The conserved secondary structure element in the unbound state contains a *base zipper motif* consisting of adenosines (Figure 2A).[38] When the tetraloop is bound, a substantial structural rearrangement occurs resulting in a so-called adenosine (or AA) platform formed by base pairing of two adjacent adenosines.[39-41] This transition is schematically illustrated in Figure 2.

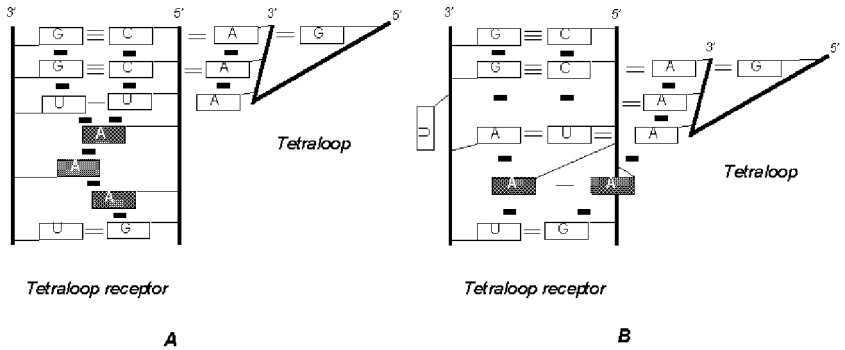


Figure 2. The GAAA tetraloop and its receptor. *A schematic illustration of the structural rearrangement of the free tetraloop receptor (A) to the bound form (B). (A) The free form of the tetraloop receptor containing the adenosine zipper motif (indicated by shaded boxes). (B) The tetraloop shifts down one nucleotide and the lower adenosine of the tetraloop now stacks on top of the adenosine platform, which is indicated by shaded boxes. Hydrogen bonds and stacking interactions are indicated by lines and solid rectangles, respectively.*[15]

A motif which was already observed in the first crystal structure for RNA is the *base triplet*.[4,5] The base triplet is plays an important role in creating and stabilising tertiary interactions.[2] Of the non-canonical base pairs, the G-U and G-A mismatch motifs are very common in RNA structures,[5,16-18,42-50] causing the RNA to adopt a distorted structure which enables a range of different types of tertiary interactions, from docking and close packing of helices [2] to tetraloop or protein binding.[15,19]

RNA structure determination

In the previous paragraphs most of the structural information was derived from biophysical data obtained by crystallographic and NMR studies. In the following section, the use of biochemical techniques to obtain structural information on RNA will be introduced.

Enzymatic structure probing - The analysis of RNA secondary structure can be accomplished via the use of RNA modifying enzymes.[51-53] A number of ribonucleases are available which are able to distinguish between double-stranded and single-stranded RNA (summarised in Table 1). Each of the enzymes listed in Table 1 cleaves the phosphate-ribose backbone of RNA, which can be monitored by using either end-labelled RNA or by primer extension.[52,53] RNases A, T1, and T2 are the most frequently employed enzymes to determine single-stranded regions in RNA molecules. RNase A cleaves single-stranded UpN and CpN containing sequences with a preference for UpA and CpA. Single-stranded GpN bonds can be detected by RNase T1 and all phosphodiester bonds in a single-stranded region are cleaved by RNase T2, irrespective of the sequence. There is only a single enzyme available for the probing of double-stranded regions in RNA: RNase V1. This ribonuclease cleaves RNA in a sequence independent manner, although at least two residues on either side of the cleavage site need to be double-stranded or stacked (*i.e.* in a helical geometry) to be a good substrate for RNase V1.[54]

As can be inferred from the molecular weights of these enzymes (Table 1), RNases are bulky molecules and therefore susceptible to steric hindrance. Hence, the information obtained from enzymatic probing experiments is very much dependent on the accessibility of the RNA backbone. Parts of the RNA which are 'buried' inside the molecule cannot be assessed by RNases. Therefore the interpretation of structure probing data obtained with RNases should be restricted to the observed cleavages and not concern the absence of cleavages. To overcome this disadvantage and acquire structural data on less accessible regions as well, chemical reagents can be used.

Chemical structure probing - Chemical reagents are much smaller than enzymes and can reach almost every part of the RNA molecule. Applicable reagents (Table 2) react with a specific chemical moiety present in the RNA. If this moiety is directly involved in any kind of interaction like for example hydrogen bond formation (e.g. in base pairing, tertiary interactions), the reactivity with the reagent will be strongly reduced or even abolished. For example, CMCT and DMS will only react with certain

nitrogen atoms on the base of nucleotides when these are not involved in hydrogen bond formation (see Table 2 and Figure 3). After modification of the base at the nitrogen atom, the modified nucleotide can easily be detected by primer extension, since the enzyme reverse transcriptase will not be able to incorporate a complementary nucleotide at the modified position and thus terminate the synthesis of the complementary strand. Alternatively, some chemical nucleotide modifications can be detected via an additional treatment of the RNA leading to strand breaks at the modified positions, which can then be monitored either by using end-labelled RNA or by primer extension. The specificity and mode of detection of the most commonly used chemical probes are summarised in Table 2 and illustrated in Figure 3.

Table 1. Enzymatic probes for RNA secondary structure determination.

Probes	MW [a]	Specificity	Phosphate [b]	Comment
RNase A	13700	UpN or GpN	3' phosphate	Minor preference for CpA and UpA
RNase T1	11000	GpN	3' phosphate	
RNase U2	12490	ApN > GpN	3' phosphate	Low pH optimum
RNase CL3	16800	CpN>>ApN>UpN	3' phosphate	
RNase T2	36000	NpN	3' phosphate	Minor preference for Ap
Nuclease S1	32000	NpN	5' phosphate	Low pH optimum; requires Zn^{2+}
RNase V1	15900	dsN or stacked	5' phosphate	Requires divalent cations

[a] Relative molecular weight.
[b] Position of phosphate at cleavage site in cleavage product.

A different kind of chemical probe is the hydroxyl radical. This highly reactive radical is believed to attack the C1 and/or C4 atom of the ribose moiety and causes subsequent cleavage of the RNA backbone.[55] It can be generated by a Fe(II)EDTA complex reacting with hydrogen peroxide in a Fenton reaction. The Fe^{2+} is oxidised to Fe^{3+}, which can subsequently be reduced by the addition of ascorbic acid.[56] The hydroxyl radical is not suitable for the determination of secondary structures, but it can be used to obtain information on the tertiary structure. A buried RNA backbone is less accessible and results in less intense cleavages by the hydroxyl radical.

Probing of RNA – protein complexes - The probing techniques described above can also be used to study RNA structures in RNA-protein complexes. In addition to footprinting the binding site of a protein, RNase probing may also identify structural changes in the RNA molecule due to protein binding. Unfortunately, RNases do not provide detailed information on the binding site of a protein due to the sterical limitations mentioned above. A very powerful technique for determining the binding site of a protein in more detail is probing the RNA-protein complex with hydroxyl radicals. Because the radical is very small, only the part of the backbone of the RNA in close contact with the protein will be less reactive.

New techniques to study structures of RNA and RNA-protein complexes - Recently some new techniques have been developed to obtain information on both

RNA and RNA-protein complexes. The development of the techniques which will be discussed below, has been made possible by recent technological advances in genetics, molecular biology, (bio)chemistry and physics.

Table 2. Chemical probes for RNA structure determination.

Probes	MW [e]	Specificity	Detection [g]	Comment
DMS [a]	126	N1-A	PE	N7-G > N1-A > N3-C; DMS also
		N3-C, N7-G	PE or EL[f]	modifies cysteine
DEPC [b]	174	N7-A	PE or EL[f]	DEPC also modifies histidine;
CMCT [c]	424	N3-U, N1-G	PE	N3-U > N1-G
Kethoxal	148	N1-G, N2-G	PE	Kethoxal also modifies the guanidino group of arginine
ENU [d]	117	Phosphates	PE or EL	Leaves ethylated phosphate on 3' oxygen
Fe(II)EDTA	17	C1' or C4' (ribose)	PE or EL	Fenton reaction

[a] dimethyl sulphate
[b] diethylpyrocarbonate
[c] 1-cyclohexyl-3(2-morpholinoethyl) carbodiimide methane-p-toluene sulfonate
[d] ethylnitrosourea
[e] relative molecular weight
[f] for detection, additional treatment of the RNA is necessary to break the backbone.
[g] PE: primer extension; EL: end-labelled RNA.

A very powerful method to detect and study important structure elements is *forced evolution*.[57] This *in vivo* technique uses viral systems with high mutation rates to search for revertants of a replication suppressive mutation. Compensatory mutations, which restore functionality of the viral system, indicate direct interactions. Functional RNAs can also be selected from a pool of RNAs which have been mutated randomly. Sequencing of the selected RNAs yields an invaluable source of phylogenetic data, helping in secondary structure reconstruction.[58] Compensatory mutations that restore structure and hence function, can also be used to test a secondary structure deduced from enzymatic and/or chemical probing. Further information can be gained from *modification interference* studies.[59,60] Nucleotides of the RNA are randomly modified with, for example, DMS or CMCT. Subsequently, functional RNAs are separated from non-functional RNAs and analysed for the difference in modification patterns indicating structurally and/or functionally important nucleotides.[61]

New methods for global assessment of the RNA structure have also been developed or improved. Angles between helices can be inferred from comparative native gel electrophoresis, exploiting the sensitivity of gel mobility to molecular structure.[62,63] A recently developed technique is *cryo-electron microscopy*,[64] which already resulted in a 23 Å structure of the entire ribosome [65,66] and in the visualisation of tRNAs on the ribosome at 20 Å resolution.[67,68] Data from crosslinking experiments can be used as a constraint in three-dimensional model building, since novel methods can introduce photoreactive nucleotides at specified positions within the RNA

molecule.[69] 4-Thio-uridine is a widely used nucleotide analogue which can be crosslinked to both RNA and proteins under the influence of UV light.[70,71] Crosslinking has been successfully used in a. o. modelling the hairpin ribozyme,[72] in characterising long range interactions in 16S RNA in the *E. coli* ribosome,[73] and in studying the dynamics of RNA-protein interactions in the HIV-1 Rev-RRE complex (using thio-guanosine).[74]

Sulphur is useful in several ways.[75] It is possible to incorporate sulphur in the backbone of the RNA, replacing the phosphorus atom.[76] Iodine can subsequently cleave the backbone at the position of the sulphur, which may be exploited in, for example, protein footprinting experiments.[77] A different application of sulphur atoms in the backbone of RNA is to determine the importance of metal ion co-ordination. As discussed above, divalent metal ions and in particular Mg^{2+}, are important in RNA folding. Sites of divalent metal ion binding can be detected by the difference in ability of metal ions to co-ordinate oxygen and sulphur. Mg^{2+} is unable to accept sulphur as a ligand, in contrast to Mn^{2+}. Thus, sulphur substitutions of oxygen that impair the effect of Mg^{2+} but can be rescued by Mn^{2+} provide evidence for a direct metal ion co-ordination site.[78,79]

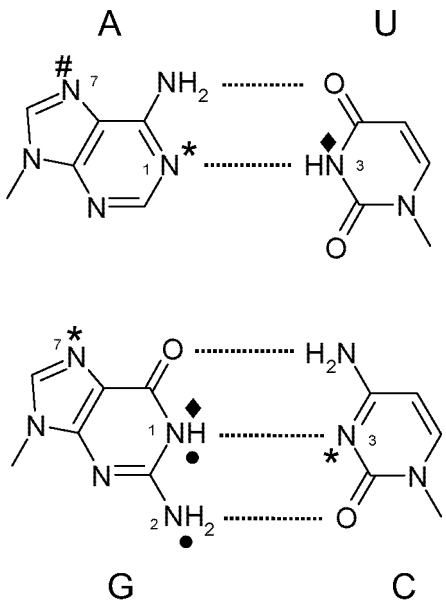


***Figure 3. Specificity of various chemical probes**. (*) DMS, (#) DEPC, (♦) CMCT, (•) kethoxal. Relevant atom numbers are indicated (see Table 2).*

A final example of a recently developed technique for global RNA structure determination is *FRET* (Fluorescence Resonance Energy Transfer). FRET relies on the principle that upon excitation of a 'donor' fluorophore, energy is transferred through space to an 'acceptor' fluorophore. The transfer of energy is dependent upon the distance between donor and acceptor, which are incorporated at two known positions. The technique can be used to accurately measure the distance between donor and

acceptor within the range of 10-80 Å.[24] Using FRET, the global structure of four-way RNA junctions and their dependency on Mg^{2+} ions has been studied.[25]

As discussed above, the hydroxyl radical attacks the ribose ring of the nucleotide and causes backbone cleavage of the RNA. There are different ways to generate this radical, depending on the chemical entity to which the iron ion is chelated. If EDTA is the ligand, the complex will be free to diffuse through the solution, probing the RNA randomly and evenly. In a recently published variant, the Fe(II) is tethered to 1-(p-bromoacetamidobenzyl)-EDTA (BABE) and can be incorporated at a specific site of the RNA, generating hydroxyl radicals from a single point.[80] Cleavage will occur only on sites within 10 Å of the tethered Fe(II) ion. The BABE complex can also be incorporated at a specific cysteine residue of a protein. RNA contacts with the protein will result in cleavages of the backbone of the RNA.[81,82] A variation of this is the attachment of (EDTA-2-aminoethyl) 2-pyridyl disulfide-Fe(III) to a specific cysteine, which has been used to study the U1A-3'UTR complex.[83] Methidium propyl-EDTA-Fe(II) can be employed to generate cleavages in double-stranded or stacked regions of RNA, due to the intercalating properties of the methidium moiety.[84] Several variations on this theme have been developed to study distinct properties of RNA and RNA-protein complexes.[75,85-87]

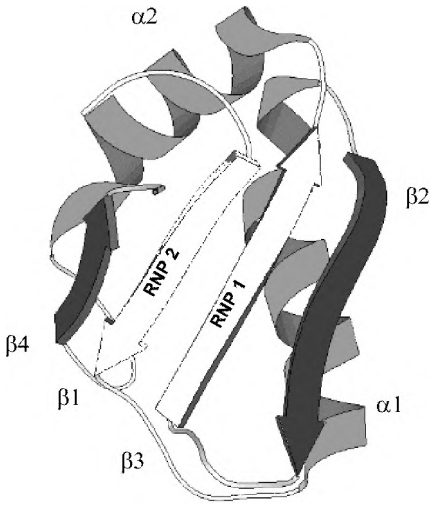


Figure 4. The RNP motif. *Tertiary folding of the RNP motif in the typical βαββαβ modular shapes. The consensus sequences RNP1 and RNP2 are indicated in the two central β strands.*[88]

RNA binding domains

The last few years have shown a steady increase in the number of RNA-protein structure determinations. Advances in isotopic labelling techniques and improved

technologies to prepare large amounts of RNA have greatly facilitated X-ray and NMR studies. These structure determinations have resulted in an increased understanding of RNA binding motifs in proteins, as well as in a much better insight in the intricate and delicate interactions between RNA and protein.

Some ten years ago, the X-ray structure of an RNA binding domain called the RNP (ribonucleoprotein) domain or RNA recognition motif: RRM was resolved (shown in Figure 4).[88] It was shown to be a four stranded antiparallel β sheet, consisting of 70-90 weakly conserved amino acids. Within this region two more highly conserved elements of eight (called RNP1) and six (called RNP2) residues, respectively, are found. The RNP1 and RNP2 elements reside in the central two strands of the β sheet (indicated in Figure 4) and play a crucial role in RNA binding.[89] The RNP motif has been found in over 200 proteins involved in RNA processing, transport, metabolism and other yet unknown functions.[90-92]

The double-stranded RNA binding motif (dsRBD) is a sequence motif consisting of approximately 65 amino acids.[93,94] It is a general double-stranded RNA-binding domain without sequence specificity, which is in contrast to the highly sequence specific RNP domain. In order to bind an RNA structure specifically, a protein can use multiple dsRBDs.[95] Similar to the RNP motif, the dsRBD folds into an antiparallel β sheet consisting of three strands.

A domain which has been found in a smaller number of proteins is the KH domain (K homology domain). In heterogeneous nuclear RNP (hnRNP) K, three copies of a short conserved sequence were found.[96-98] Although the natural target has not yet been identified, minor modifications in the sequences resulted in the loss of binding to poly-(U) *in vitro*.[99] Very recently a new sequence motif was discovered in the Sm proteins, which form the core domain of small nuclear ribonucleoprotein particles (snRNPs) involved in pre-mRNA splicing.[81,100,101] The crystal structures of two Sm protein complexes (D3B and D1D2) have been solved and they suggest that the seven Sm proteins could form a closed ring. The hole of the ring could be occupied by the RNA.[102]

Introduction to ribonucleoprotein particles

Autoimmunity

Patients suffering from an autoimmune disease often generate antibodies against self-antigens. The cause for this loss of ability to distinguish between self and non-self by the immune system (breaking of tolerance) is still unsolved. Most autoantibodies are directed against RNA-protein particles, called ribonucleoprotein particles (RNPs). Most frequently the proteins in the RNP complex are targeted although in some cases the isolated RNA moiety can also be recognised by a subset of the autoantibodies. Sera of autoimmune patients can be very helpful in the isolation and study of autoantigens. Knowledge of their structure and function might help to solve the aetiology of these autoimmune disorders.

U1 snRNP

The U1 snRNP complex consists of one RNA molecule (U1 snRNA) bound by several proteins. Three proteins (U1A, U1C and the U1-70K protein) are specifically associated with U1 snRNP. The remaining U1 snRNP proteins (Sm proteins) are associated with other snRNPs as well (for a review see Klein Gunnewiek et al., 1997).[151] The U1 snRNP is often recognised by autoantibodies from patients with systemic lupus erythematosus (SLE). In fact, serum from patients with SLE was used to demonstrate for the first time that some snRNP-associated proteins were contained in different snRNPs. U1 snRNP plays, together with other U snRNPs, an important role in the splicing of pre-mRNA. It recognises and binds the 5' splice site after which U2 and U4/U5/U6 join the complex. This large spliceosomal complex subsequently performs the excision of the intron and the subsequent ligation of the exon sequences.

The secondary structure of the U1 snRNA is shown in Figure 5A. It consists of five stem-loop structures and a four-way junction. Stem-loop I is bound by the 70K protein and stem-loop II by the U1A protein. The Sm proteins are bound to the single-stranded region between III and IV (Figure 5A). Sera from SLE patients may not only contain antibodies directed to the protein components of the U1 snRNP, but also antibodies directed to the U1 snRNA itself.[152] Previous studies have shown that the main targets of these specific anti-U1 snRNA autoantibodies are the stem of stem-loop II and the loop of stem-loop IV (Figure 5A). Both targets are regions that are not known to be associated with protein. U1A binds primarily to the loop of stem-loop II (*i.e.* the AUUGCAC sequence)and for stem-loop IV no binding protein has yet been identified.

The U1A protein consists of 282 amino acids (34 kDa) and contains two RNP motifs. The N-terminal RNP motif of U1A is known to be able to bind two RNA targets (Figure 5), while for the C-terminal domain no RNA counterpart has been identified yet.[103] The primary target of the N-terminal RNP motif is the loop region of stem-loop II of U1 snRNA.[104-110] The three-dimensional structure of this RNP motif was studied by X-ray crystallography and NMR [88,111] and is shown in Figure 4.

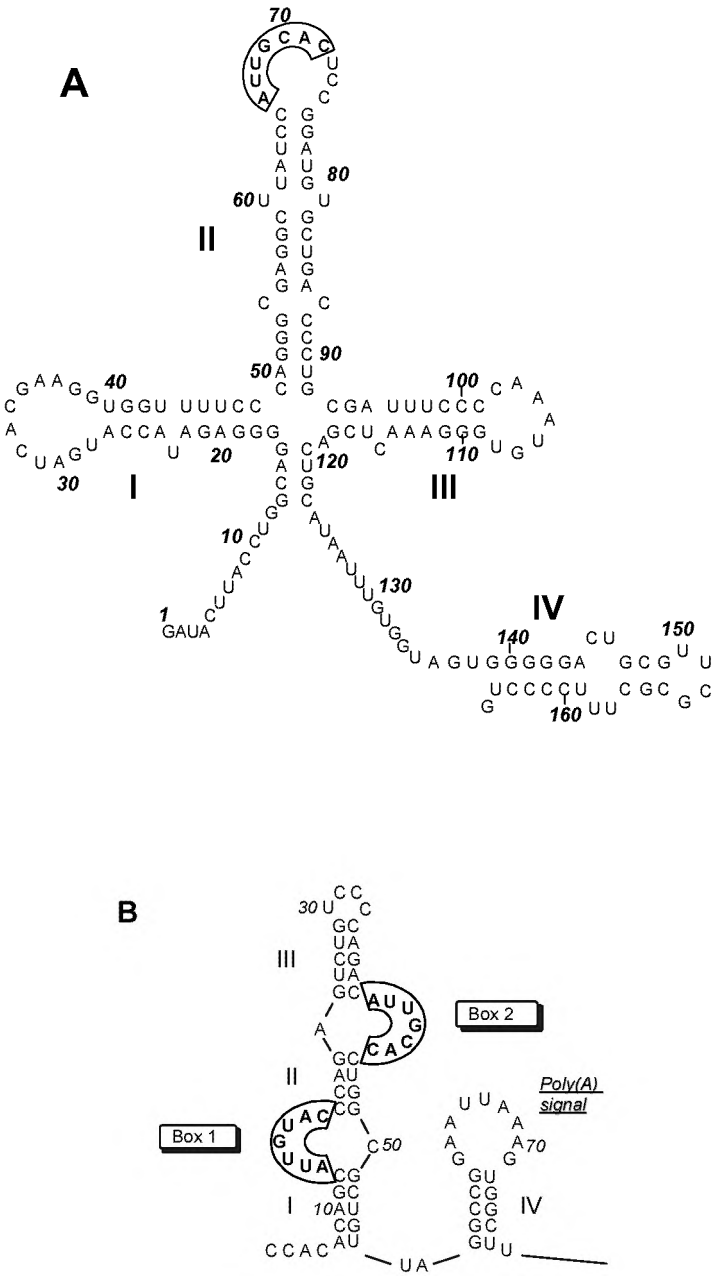


***Figure 5. RNA targets of the N-terminal RNP motif of U1A**. (A) Secondary structure of U1 snRNA. The stem structures are numbered I to IV. (B) The 3'UTR of U1A pre-mRNA. Box 1, box 2, the poly(A) signal and the stem structures are indicated. The U1A-binding sequences are boxed in both RNAs.*

Stem-loop II of U1 snRNA contains a 7-nucleotide recognition sequence AUUGCAC within the 10 nucleotide loop (Figure 5A). The structure of the N-terminal RNP motif of U1A complexed to stem-loop II has been investigated by NMR [112-114] and cross-linking studies.[115] These studies showed that the β-sheet provides an excellent surface for RNA binding.

A few years ago, the crystal structure of the complex has been determined as well, showing the delicate interactions between the U1 RNA and the protein.[89,90,116] Two lysine residues co-ordinate to the phosphate backbone of the RNA helix. Together with the arginine at position 52, they position the RNP motif on the RNA. Upon binding, a third alpha-helix of the U1A protein (extending from β4, Figure 4) is shifted from the middle of the β-sheet to the top of the sheet (in the direction of the top of the RNP motif as represented in Figure 4), where it is interacting with the RNA.[117] This is an example of *induced fit*, that is: both the protein and the RNA rearrange their tertiary structures to accommodate binding.[113,118] Some of the additional specific interactions are achieved by stacking of bases on tyrosine and phenylalanine residues and by hydrogen bonding between nucleotides and the protein backbone.[89,115] The importance of tyrosines (and in particular tyrosine-13, positioned in RNP2) in this RNA-protein interaction was analysed by NMR using point mutations.[119] Not only has tyrosin-13 stacking interactions with C70 (Figure 5A),[89] it also has interactions with glutamine-54 which has in turn an important role in stabilising the backbone geometry, forming a network of interactions that control the association with the RNA.[119]

The expression level of the U1A protein is very elegantly regulated. The primary binding site for the U1A protein is located on the U1 snRNA, as described above. At relatively high U1A protein concentrations a second target can be bound: the 3'UTR of newly synthesised U1A pre-mRNA. When two copies of the protein are bound to this 3'UTR, polyadenylation of this same pre-mRNA is inhibited.[120,121] This results in destabilisation and degradation of the U1A pre-mRNA, causing a reduced production of U1A protein. Thus, U1A protein is able to autoregulate its own expression level *in vivo* (Figure 6).[121-123] The 3'UTR of the pre-mRNA contains two binding sequences for the U1A protein, one identical to the sequence in the U1 snRNA and the second one containing a single nucleotide substitution (AUUCUAG). These sequences are called Box 2 and Box 1, respectively (shown in Figure 5B).[120] Box 2 binds the U1A protein with the same affinity as U1 snRNA, whereas Box 1 binds U1A ~30 fold less strongly.[121] The secondary structure of the 3'UTR has been investigated using enzymatic probing [121] and by NMR[124] and is shown in Figure 5B. Box 1 and Box 2 are on opposite sides of the central helix. In a theoretical three-dimensional folding, the two boxes are positioned on the same side of the molecule,[125] acting like a pair of hands folded around the two RNP motifs.

The Ro RNPs

Ro RNPs are small cytoplasmic particles consisting of one of four Y RNAs (cYtoplasmic RNAs: Y1, Y3, Y4, and Y5) and at least two proteins, designated Ro60 and La. Ro RNPs most likely contain other, yet unidentified proteins as well. Although the existence of Ro RNPs has been discovered quite some time ago, their function in the cell is still not known. Recently, however, a synergistic role for the Ro60 and La proteins in regulating the translation of L4 ribosomal protein mRNA has been suggested in *Xenopus laevis.*[126] Most interestingly, an as yet unidentified RNA was found to be

associated with this complex, suggesting that the whole Ro RNP complex might be involved.[126]

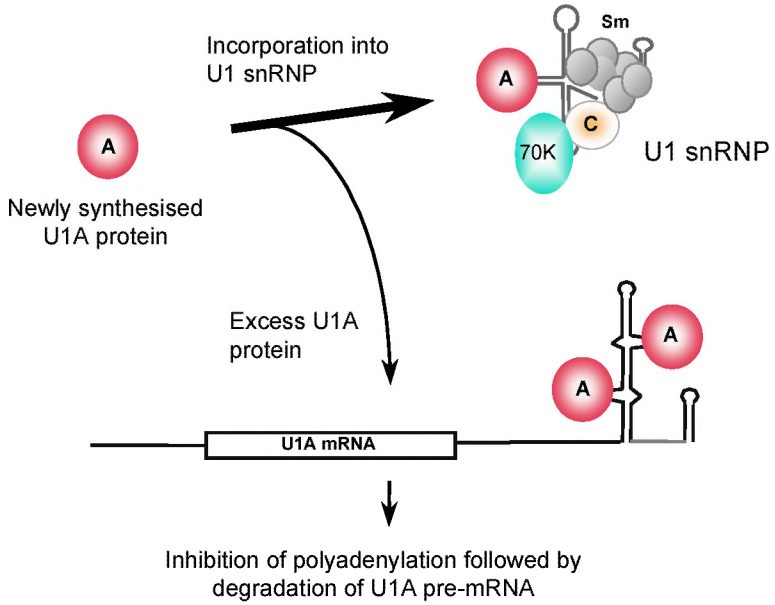


Figure 6. Autoregulation of U1A protein expression. *Newly synthesised protein is bound to the U1 snRNP. If there is an excess of U1A protein, compared to the level of U1 snRNP, the excess protein may bind its own pre-mRNA.*[120]

The Ro60 protein has a molecular weight of 60 kD,[127,128] and contains one RNP domain (Figure 7). [129] It is the major protein component of Ro RNPs and binds to the lower part of the Y RNAs in the region of the bulged C9 [127,130] which is important for binding (see Figure 8).[131] Ro60 also contains a zinc finger motif, but this element is not conserved in the *Xenopus laevis* or *Caenorhabditis elegans* counterparts. Deletion studies showed that Ro60 needs almost all of its amino acids in order to bind the RNA.[131] In *Xenopus laevis*, the Ro60 homologue is also found complexed with variants of pre-5S rRNAs.[132] Because these non-functional 5S rRNA precursors are processed inefficiently and eventually are degraded, Ro60 has been proposed to function in a quality control pathway for 5S rRNA biosynthesis.[132]

The La protein has been found associated with many RNA targets and hence several functions have been proposed for this protein (Figure 7). La is a 47 kD protein, containing two RNP motifs and plays an important role in RNA polymerase III transcription termination. It is transiently associated with all RNA polymerase III transcripts, with the exception of the Y RNAs, for which the association is not transient. Recently, it has been identified as an RNA polymerase III (re)initiation factor as well (reviewed by Pruijn *et al.*).[133] The La protein binds Y RNAs at the oligo-(U) stretch at the 3' end of the RNA (Figure 8).

The human Y RNAs range in size from 84 to 112 nts. and are transcribed by RNA polymerase III.[134,135] The secondary structures of hY1 and hY5 RNA have been investigated in detail [136] and are shown in Figure 8 together with the proposed

secondary structures for hY3 and hY4 RNA. The general secondary structure of Y RNAs consists of three or four stem structures separated by (asymmetric) internal loops. Y RNA sequences from a number of mammals, *Xenopus laevis*, *Iguana iguana* and *Caenorhabditis elegans* have been determined (reviewed by Pruijn *et al.*).[133] The proposed secondary structures are supported by these phylogenetic data, which also indicate that Y3 is evolutionarily the most conserved of the Y RNAs.[137]

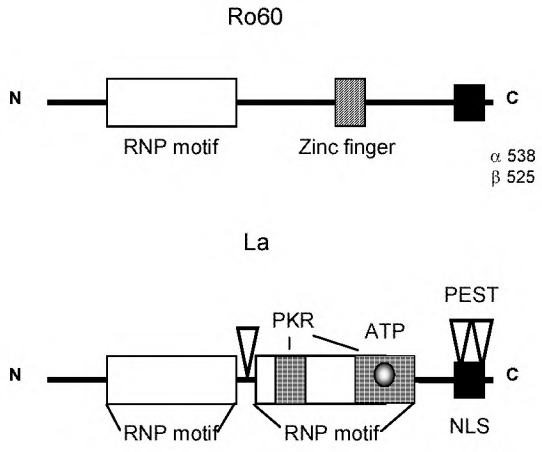


Figure 7. *Schematic representation of the Ro60 and La proteins. Structural motifs are indicated: the RNP motif, Zinc finger domain, PKR homology domain, putative ATP binding site, nuclear location signal (NLS) and PEST regions (regions susceptible to proteolytic cleavage). For Ro60 two splicing variants (α and β) are known.*

Outline of this thesis

The main theme of this thesis is to gain knowledge about the structure of RNAs and their interaction with proteins in autoantigenic RNPs. Better understanding of the structure of such complexes may lead to more insight into their cellular function and help to answer the question why they are targeted by autoantibodies of patients with an autoimmune disease.

Chapter 2 describes a structural analysis of the 3'UTR of the U1A pre-mRNA, both in the absence and in the presence of the U1A protein. Secondary structure probing experiments on the naked RNA were performed using chemicals such as described in Table 2. The RNA-protein complex was studied using enzymes and Fe(II)EDTA. In part these experiments were a continuation of the work performed by Van Gelder et al.,[121] resulting in a more detailed view on the structure of the RNA and on the dynamics of the RNA upon protein binding.

The availability of murine monoclonal antibodies to study the U1 snRNP particle is limited to the protein components of the complex. However, to study the RNA moiety

of the particle one needs specific monoclonal anti-RNA antibodies. The phage display technique was successfully used to generate monoclonal antibodies against the U1 snRNA. In Chapter 3 a recombinant autoantibody fragment (scFv) directed to the U1 snRNA was characterised and the binding site determined using enzymatic probing and Fe(II)EDTA.

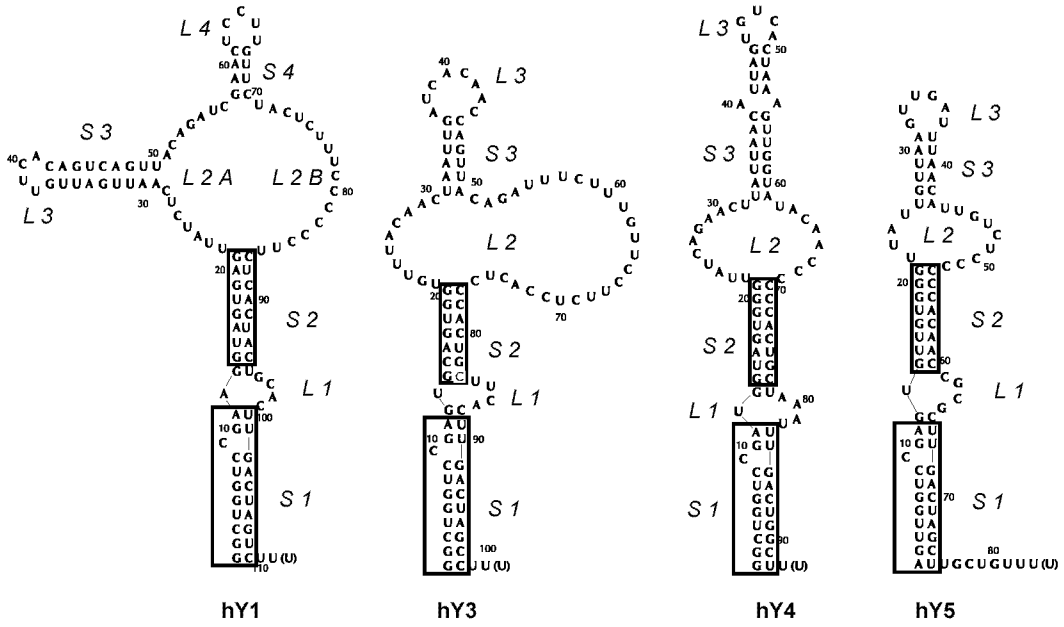


***Figure 8. The secondary structure of human Y RNAs**. The names of the different stem and loop structures are indicated.*

Of the four human Y RNAs, the structures of Y1 and Y5 RNA have been most extensively studied.[136] Chapter 4 reports experimental data on the structures of the other two Y RNAs, *i.e.* hY3 and hY4. Furthermore, the RNA structures of frog and iguana Y3 and Y4 were determined in order to unravel conserved structural elements in these RNAs. Enzymatic and chemical structure probing experiments were performed and combined with computer generated secondary structures to deduce the most stable solutions structure(s).

Chapter 5 contains a general discussion of the work presented in this thesis.

2

Probing the 3' UTR structure of U1A mRNA and footprinting analysis of its complex with U1A protein

The structure of the conserved region of the U1A pre-mRNA (Ag RNA) and its complex with U1A protein was investigated. The previously proposed secondary structure of Ag RNA, derived from enzymatic probing and analysis of structure and function of mutant mRNAs, is now confirmed by chemical probing data and further refined in the regions where the enzymatic data were not conclusive. The two unpaired nucleotides in the internal loops opposite of the Box sequences as well as the tetraloop could not be cleaved by ribonucleases, but are accessible to chemical probes. Concerning the RNA-protein complex, the protection experiments showed that the Box regions are largely protected when the U1A protein is present. All stem regions in the 5' part of the structure seem protected against ribonucleases. Unexpectedly, the nucleotides of the tetraloop become accessible to ribonucleases in the RNA-protein complex. This result indicates that the tetraloop undergoes a conformational change upon U1A protein binding. The 3' part of the Ag RNA sequence, containing the polyadenylation signal in a hairpin structure, showed hardly any protection, a finding that agrees with the fact that U1A does not interfere with the binding of the cleavage polyadenylation specificity factor (CPSF) to the polyadenylation signal.

S. W. M. Teunissen, C. W. G. van Gelder, W. J. van Venrooij.
Biochemistry, ***36****, 1782-1789 (1997)*

Introduction

The removal of introns from pre-messenger RNA, known as splicing, is an important process in which several small ribonucleoprotein particles (snRNPs) participate. One of them, U1 snRNP, contains a U1 snRNA molecule, at least eight Sm proteins also present in other U snRNPs, and three U1-specific proteins named U1-70K, U1C and U1A (reviewed by Lührmann et al., 1990).[138]

The U1A protein contains two copies of an RNA binding domain, also referred to as RRM (RNA Recognition Motif), RNP80 motif or RNP motif, one at the N-terminal and one at the C-terminal region of the protein. The N-terminal RNP motif binds directly to the second stem-loop of U1 snRNA.[88,104,106,114,139] The structure of this RNA-binding domain of the U1A protein has been determined by X-ray crystallography and NMR studies [88,111] and consists of a four stranded antiparallel ß-sheet with two α-helices both lying on the same side of the sheet. The loop of the second hairpin of human U1 snRNA contains 10 nucleotides. It has been shown that the first seven of them (with the highly conserved sequence AUUGCAC), are critical for U1A binding, whereas the structural context of this sequence affects binding affinity.[104,140-142] Recently, the complex between the N-terminal RNP motif of U1A and the second stem-loop of U1 snRNA has been investigated by NMR [112,113] and cross-linking studies.[115] The ß-sheet of U1A was shown to form the recognition surface, where the main contacts with the loop of the U1 snRNA hairpin occur. Furthermore, the crystal structure of this RNA-protein complex has been determined,[89] revealing detailed information on the interaction between U1 snRNA and the U1A protein.

It has been shown that the 3' UTR of the U1A pre-mRNA (Ag RNA) contains a region which has been conserved among vertebrates.[120] This region contains two stretches of seven nucleotides (called Box 1 and Box 2) which have a sequence similar to that contained in the second stem-loop of U1 snRNA and which are located in close proximity to the polyadenylation signal. Two human U1A proteins can bind to these two Box regions [120,121] and experiments *in vitro* and *in vivo* have shown that the binding of two U1A proteins to this region specifically inhibits polyadenylation of the U1A pre-mRNA.[120] Thus, the U1A protein regulates the metabolism of its own pre-mRNA. The mechanism of this regulation has been elucidated by *in vitro* studies in which the U1A protein was shown to inhibit both specific and non-specific polyadenylation by poly(A) polymerase (PAP) via a specific interaction in which the C-termini of both proteins seem to be involved.[122]

Recently the human U1A protein - Ag RNA complex and the relationship between its structure and function in inhibition of polyadenylation *in vitro* was investigated.[121] The secondary structure of the conserved region of the 3' UTR was determined via a combination of theoretical predictions, phylogenetic sequence alignment, enzymatic structure probing and analysis of structure and function of mutant mRNAs. It was shown that the integrity of much of this secondary structure is

required for both high affinity binding to U1A protein and specific inhibition of polyadenylation *in vitro*.[121]

Here a more detailed analysis of the Ag RNA and its complex with U1A protein is reported. Chemical structure probing was performed on the Ag RNA both at 20 °C and at 0 °C, which provided a better understanding of some structural features which were not well understood from the enzymatic probing experiments. Both the reactivity of atoms involved in Watson-Crick base pairing and the reactivity of the N7-atoms of purines was investigated. Furthermore, the U1A - Ag RNA complex was probed by Fe(II)EDTA and several enzymes, yielding not only a footprinting pattern, but also structural information on the complex.

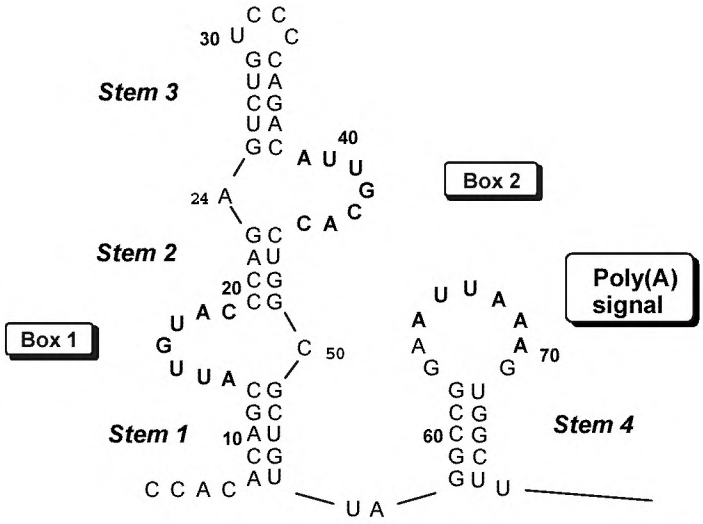


Figure 1. Secondary structure of Ag RNA. *Proposed secondary structure and nomenclature of the conserved region of the 3' UTR of the human U1A pre-mRNA (Ag RNA). The Box sequences, the polyadenylation signal and stems 1, 2, 3 and 4 are indicated.*

RESULTS

Structure probing of Ag RNA

The proposed secondary structure of the conserved region of the 3' UTR of the U1A mRNA, called Ag RNA, is shown in Figure 1 and consists of two distinct parts which are separated by only two nucleotides, U56 and A57.[121] The 5' part, which has a symmetric structure, contains three stems (numbered 1, 2, and 3), separated by two asymmetrical internal loops containing the Box 1 and 2 sequences, which are required

for U1A protein binding.[120] The 3' part of the structure is formed by stem 4 and a 9-nucleotide loop containing the polyadenylation signal (loop 4).

Enzymatic structure probing experiments showed that the central three nucleotides in Box 1 and 2, as well as the polyadenylation signal were single-stranded.[121] The presence of the highly conserved stems 2 and 3, which are needed for U1A protein binding, was established by RNase V1 cleavages and by analyses of structure and function of mutant mRNAs.[121] However, the behavior of a few other parts of the structure was less easy to understand. Stems 1 and 4 showed cleavage both by RNase V1 and by single-strand-specific ribonucleases. This seemed to indicate that these two stems, which have not been strongly conserved in evolution and which appear not to be important for either U1A protein binding or inhibition of polyadenylation by the U1A protein, are of weak stability or may not exist at all in solution.[121] Furthermore, the tetraloop of stem 3 (nucleotides 30-33) was hardly cleaved by ribonucleases under native conditions, suggesting that its structure might be very compact. A similar behavior was found for the unpaired nucleotides A24 and C50. This could arise either from the possibility that these two nucleotides are located inside the helix or from the possibility that other parts of the Ag RNA sterically hinder the ribonucleases. In contrast, chemical probes are, due to their small size, less sensitive to steric hindrance and therefore could provide more detailed insight in the Ag RNA structure.

The four bases were monitored at their Watson-Crick base-pairing positions by DMS at N1-A and N3-C and by CMCT at N3-U and N1-G. Position N7 of guanine and adenine residues was probed by DMS and DEPC, respectively. The experiments were performed under native, semi-denaturing, and denaturing conditions. Tertiary interactions are generally less stable than Watson-Crick interactions and are expected to melt under semi-denaturing conditions.[52] Experiments under such conditions will also give information about the stability of the different double stranded domains. Ag RNA was probed both at 20 °C and at 0 °C. The latter temperature was used to minimize the breathing in this relatively small RNA molecule, a phenomenon observed at 20 °C (see below).

Figures 2A through 2C show examples of the chemical probing results for Ag RNA, while Figure 3 summarises the results of several independent probing experiments for both 20 °C and 0 °C, for the Watson-Crick base pairing positions (Figure 3A) and for the N7-atom of purines (Figure 3B).

Stem regions

Stems 2 and 3. At 20 °C the presence of stems 2 and 3 is clearly supported by the chemical modification data, since many nucleotides are only reactive under denaturing conditions. This is shown, for example, for nucleotides U26 and U28 in stem 3 in Figure 2A (lane 9) for the CMCT reaction. Their counterparts in stem 3, A37 and A35, respectively, are reactive with DMS (data not shown), as is A22 in stem 2 (see Figure 2B, lane 3). This difference in reactivity between adenosines and uridines in A-U

base pairs has also been observed in helical regions of other RNAs [52,136,145] and is probably related to the fact that DMS (M_r=126) is smaller than CMCT (M_r=424).

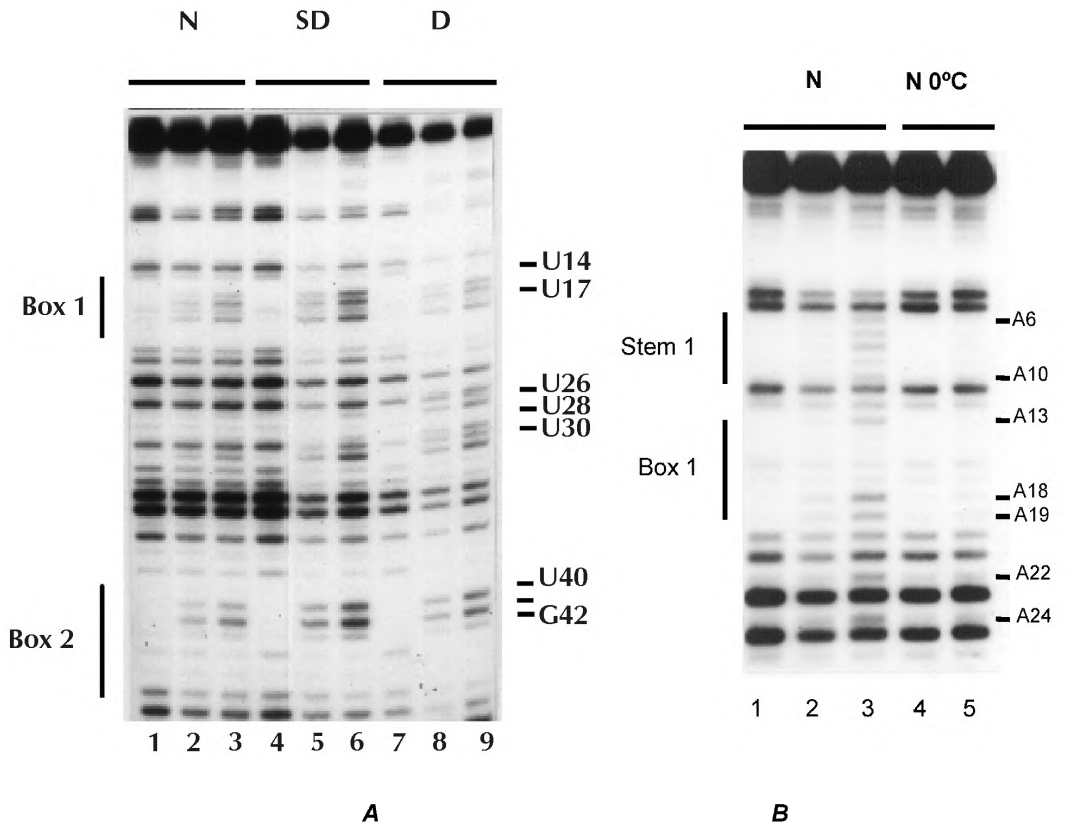


Figure 2. Secondary structure probing of Ag RNA. *(A)Chemical probing of Ag RNA with CMCT. Detection of modifications was achieved via primer extension. Samples in lanes 1, 4 and 7 are control incubations in which reagent was omitted. Lanes 2, 3, 5 and 6: 50 µl CMCT (final concentration: 0.33 M) incubated at 20 °C for 20, 30, 5 and 10 min, respectively. Lanes 8 and 9: 10 and 25 µl CMCT (final concentration: 0.15 M), incubated for 1 min at 90 °C. (B)Chemical probing of Ag RNA with DMS at 20 °C and at 0 °C. Detection of modifications was achieved via primer extension. Samples in lanes 1 and 4 are control incubations in which reagent was omitted. The reaction conditions are indicated: N (native conditions, 20 °C) and N 0 °C (native conditions, 0 °C). Lane 2: 0.5 µl DMS incubated for 15 min (final concentration: 0.1 x 10^{-9} M). Lanes 3 and 5: 1.5 µl DMS (final concentration: 0.3 x 10^{-9} M) incubated for 15 min.*

Concerning the N7 positions of the purines in stems 2 and 3, the guanosines are on the average more reactive towards DMS than the adenosines towards DEPC (see Figure 3B for a summary). This difference occurs because DEPC is larger than DMS, and therefore more sensitive to stacking,[146,147] which in our case is most clearly visible in the experiments performed at 0 °C. Most nucleotides show reduced accessibility at 0 °C and two N7 atoms, A35 and A37, cannot be modified at all at this

temperature (Figure 2C, lanes 3 and 6). Guanosines 23, 25, 49 and 51 are bordering the two internal loops where they are likely to be more accessible, a behaviour also found in other RNA internal loops.[136]

Stems 1 and 4. In agreement with the enzymatic probing,[121] the chemical probing experiments showed that stems 1 and 4 are of weak stability and are breathing at 20 °C. In stem 1, many nucleotides were reactive at their Watson-Crick positions at this temperature (see Figure 2B, lane 3, for nucleotides 6-10). When we lowered the

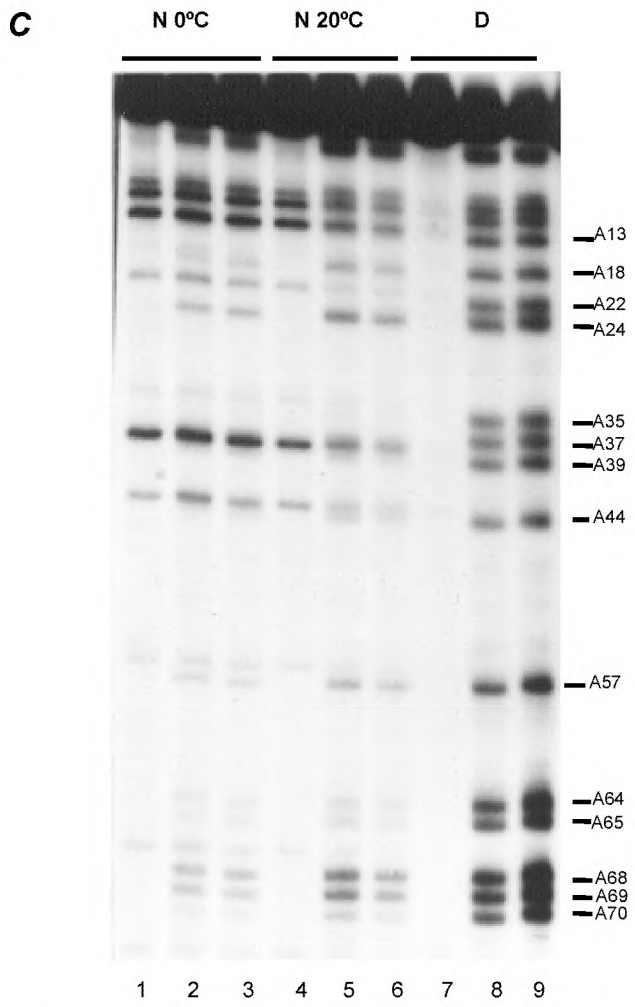


Figure 2 (cont.) *(C)Chemical probing of 3'-end-labelled Ag RNA of N7-A positions with DEPC. Samples in lanes 1, 4, 7 are control incubations in which reagent was omitted. The reaction conditions are indicated above the figures: N (native conditions), D (denaturing conditions). Lanes 2 and 3: 10 and 20 µl DEPC (final concentration: 0.3 x 10^{-9} and 0.6 x 10^{-9} M, respectively) incubated for 1 hr at 0 °C. Lanes 5 and 6: 10 and 20 µl DEPC (final concentration: 0.3 x 10^{-9} and 0.6 x 10^{-9} M, respectively) incubated for 50 min at 20 °C. Lanes 8 and 9: 2 and 4 µl DEPC (final concentration: 0.06 x 10^{-9} and 0.12 x 10^{-9} M, respectively) incubated for 4 min at 90 °C.*

temperature to 0 °C, the Watson-Crick positions of nucleotides 6-10 could no longer be modified by DMS (Figure 2B, lane 5). Only U55 could still be modified by CMCT at this temperature (data not shown), but since this nucleotide is at the end of stem 1 it is likely to be more reactive. The reactivity of the N7-atoms showed a similar temperature dependent behaviour. The N7-atoms of G51 and G54 were reactive at 20 °C, while their

reactivity was reduced at 0 °C (data not shown). In stem 4 several nucleotides showed reactivity at 20 °C, both at their Watson-Crick and N7-positions (see Figure 3 for a summary), which was reduced when the temperature was lowered. Taken together, these results suggest that stem 1 and 4 indeed can be formed, although they are of weak stability at 20 °C.

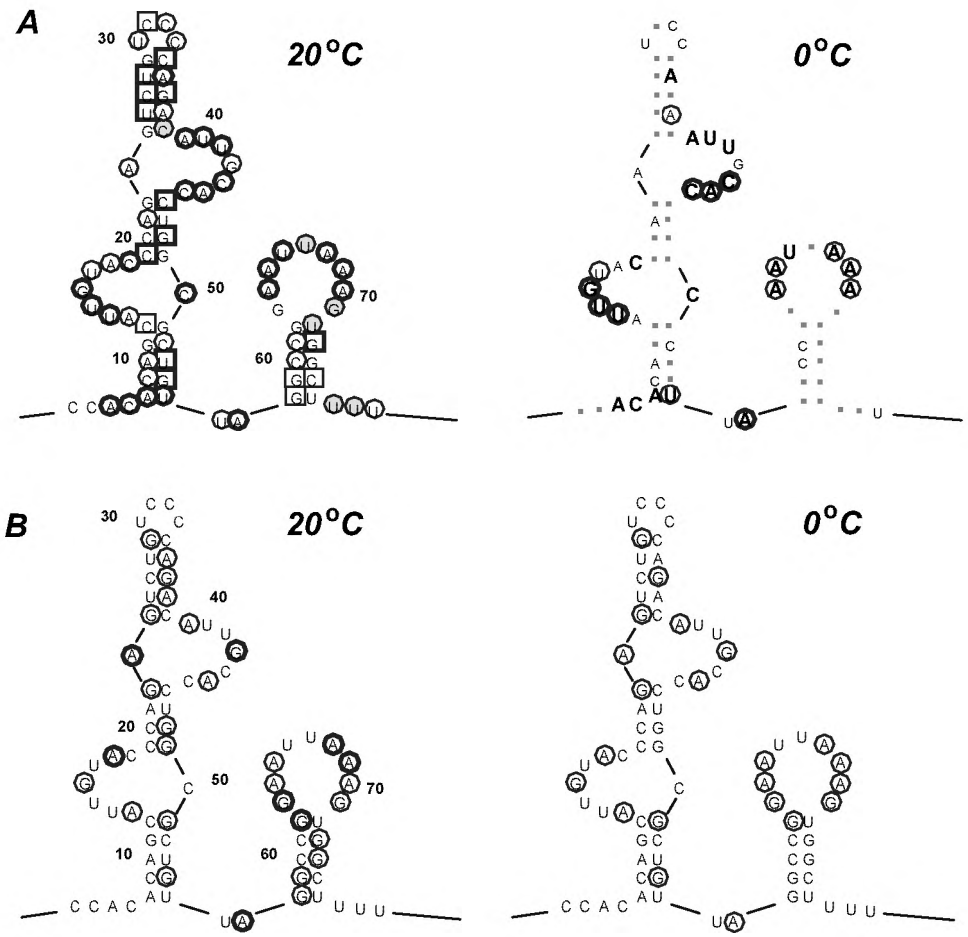


Figure 3. Summary of the chemical structure probing of Ag RNA. *(A)Reactivity of Watson-Crick positions of nucleotides in Ag RNA towards chemical probes at 20 °C and 0 °C. Consensus data from several independent experiments using both primer extension and end-labeled detection are shown.* ***Left panel****: results from experiments performed at 20 °C. Reactivity under native conditions is indicated by circles. The thickness of the circles represents moderate (thin) or strong (thick) reactivity. Reactivity under semi-denaturing conditions is indicated by filled circles. Squares indicate reactivity under denaturing conditions. The thickness of the squares represents moderate (thin) or strong (thick) reactivity.* ***Right panel****: results from experiments performed at 0 °C; nucleotides which were not reactive at 20 °C are indicated by shaded boxes; the nucleotides which were reactive at 20 °C are indicated in standard and bold font (moderate and strong reactivity, respectively). Reactivity at 0 °C is indicated by circles, as in the left panel. Nucleotides for which no reactivity is indicated show RT stops in the primer extension reaction. (B)Reactivity of N7-positions of purines in Ag RNA towards chemical probes at 20 °C (left panel) and 0 °C (right panel). Consensus data from several independent experiments using 3'-end-labelled Ag RNA are shown. Symbols are identical to those used in Figure 3A. Data are available for nts. 11-79.*

Loop and linker regions

Box 1 and Box 2 regions. All nucleotides in the Box 1 and 2 sequences are accessible at their Watson-Crick positions at 20 °C (see Figure 3A for a summary of the data). Figure 2B (lane 3) shows accessibility of A13, A18 and C19 in Box 1 to DMS and Figure 2A (lane 3) shows the accessibility of U14 to U17 in Box 1 and of U40 to G42 in Box 2 to CMCT. At 0 °C a few nucleotides at the 5' part of Box 2 become inaccessible at their Watson-Crick positions (Figure 3A, right panel). This is shown for the N1 atoms of A13, A18, and C19 in Figure 2B, lane 5. This behaviour is probably due to stacking of the bases which could agree well with the RNase V1 cleavage found at the 5' parts of both Box sequences.[121] It must be noted, however, that both Box sequences do not behave exactly the same at 0 °C (see Discussion).

Although RNases were unable to cleave the unpaired nucleotides A24 and C50, our chemical probing results show that N3 of C50 is strongly reactive (not shown) and that the N1 atom of A24 (Figure 2B, lane 3) is moderately reactive at 20 °C. For C50 this had to be deduced from reactions with 3'-end-labelled RNA (data not shown) due to the occurrence of a natural stop of reverse transcriptase at C50 in the primer extension reactions. Note that for A24 also the N7 atom is available for modification (Figure 2C, lane 6). At 0 °C, N1 of A24 is no longer accessible (Figure 2B, lane 5), probably due to stacking, but the N7-atom of A24 still can be modified (Figure 2C, lane 3).

Tetraloop. In the tetraloop U30 is moderately reactive towards CMCT while the reactivity of cytosines 31-33 toward DMS is more difficult to evaluate due to the presence of natural RT-stops, especially at nucleotides 33 and 34. By using 3'-end-labelled RNA, however, it was found that nucleotides 32 and 33 are moderately reactive at their N3 position, while N3 of C31 was only reactive under denaturing conditions (data not shown).

Loop 4 and linker region. In full agreement with the enzymatic probing,[121] loop 4 was completely accessible at the Watson-Crick positions under native conditions with the exception of U67, which became accessible to CMCT only under SD conditions (data not shown). The N7 positions of the purines (see Figure 2C, lane 6) are accessible and A68 through A70 seem more strongly modified than A64 and A65. At 0 °C most of the nucleotides were still moderately accessible at both Watson-Crick and N7 positions, but the reactivity was clearly reduced compared to 20 °C (shown in Figure 2C, lanes 3 and 6).

The linker region, which connects the 5' part with the 3' part of the structure, is formed by the two nucleotides U56 and A57. Both nucleotides are fully accessible at 20 °C (data not shown). At 0 °C, U56 can no longer be modified but A57 is still accessible, both at N7 (Figure 2C, lanes 2 and 3) and at N1 (see Figure 3A).

Analysis of the complex of Ag RNA with U1A protein

To obtain information on the complex of Ag RNA and the U1A protein, footprinting experiments were performed using various ribonucleases and Fe(II)EDTA. In these experiments 5'-end-labelled Ag RNA was incubated with an excess of U1A protein. After this, the resulting RNP complexes were probed with RNases A, T1, T2, V1,

or with hydroxyl radicals. Examples of RNase footprinting are shown in Figure 4A, Figure 4C (both RNase T2) and Figure 4B (RNase V1), while Figure 4D summarizes the results obtained by ribonuclease protection.

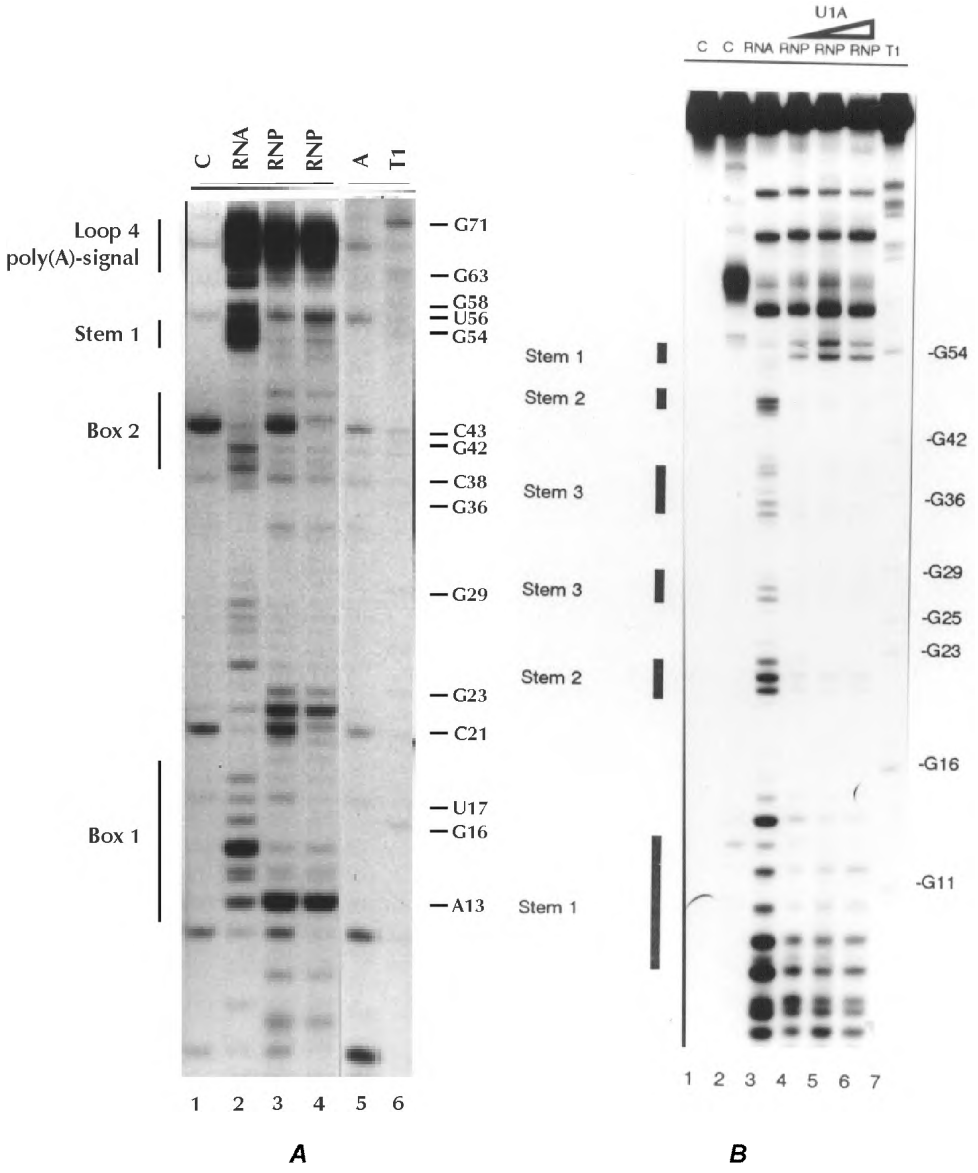


***Figure 4. Enzymatic probing of the U1A-U1A pre-mRNA complex**. (A) Enzymatic footprinting using 5'-end-labelled Ag RNA and single-stranded-specific RNase T2. Lane 1: Control incubation where U1A protein and RNase T2 were omitted; Lane 2: RNA probed at 20 °C for 10 min with RNase T2 (5 x 10^{-3} U); Lanes 3 and 4: RNA incubated with respectively 150 and 300 molar excess of U1A protein, probed with RNase T2 (5 x 10^{-3} U); Lanes 5 and 6: RNA probed under denaturing conditions with RNases A and T1, to obtain a sequence ladder for U/C and G, respectively. (B) Enzymatic footprinting using 5'-end-labelled Ag RNA and RNase V1. Lane 1: Control incubation where U1A protein and RNase V1 are omitted; Lane 2: Control incubation where both Ag RNA and U1A protein are present, but where RNase V1 is omitted; Lane 3: RNA probed at 20 °C for 10 min with RNase V1 (0.06 U); Lanes 4, 5 and 6: RNA incubated with respectively 150, 300 and 500 molar excess of U1A protein, probed with RNase V1; Lane 7: RNA probed under denaturing conditions with RNase T1, to obtain a sequence ladder.*

As might be expected, the Box 1 and 2 regions are almost completely protected by the U1A protein (compare lanes 2 and 3 in Figure 4A). The phosphodiester bond between C43 and A44 is very sensitive in both RNA and RNP, which obscures clear interpretation of the protection pattern at that position. Such intrinsic fragility is a well known phenomenon in RNA, especially for pyrimidine-adenosine bonds.[147] Note that in the experiment shown in Figure 4A, probably some RNase A-like activity was present (Figure 4A, lane 1). A13 in Box 1 becomes a hypersensitive site in the RNP complex (Figure 4A, lane 3). The single-strand-specific RNases also cleave some nucleotides in the stem regions in the naked RNA, for instance nucleotides 26-28 in stem 3 and nucleotides in both stem halves of stem 1. These cleavages are absent or much weaker in the RNP (Figure 4A, compare lanes 2 and 3; see also Discussion).

The protection pattern obtained by RNase V1 in the presence of U1A wt protein (Figure 4B) shows that the stem regions in the 5' part of the RNA (stems 1, 2 and 3) are protected in the RNA-protein complex, while the 3' part remains unprotected (compare lanes 3 and 4). One V1 cleavage, between nucleotides U53 and G54, became stronger in the RNP as compared to the naked RNA. Footprinting experiments with RNase V1 and T2 in the presence of $U1A_{101}$ (containing the N-terminal 101 amino acids of U1A) did not show significant differences in the protection pattern (data not shown).

Because nucleotides of the tetraloop in the naked RNA were not cleaved by RNases (see above [121]) information about protection of this region was not expected to be obtained. Interestingly, however, the tetraloop seems to become accessible in the RNP complex, indicating a structural change in this part of the RNA upon protein binding (see Discussion). This is shown in Figure 4C, lanes 4-6, where RNase T2

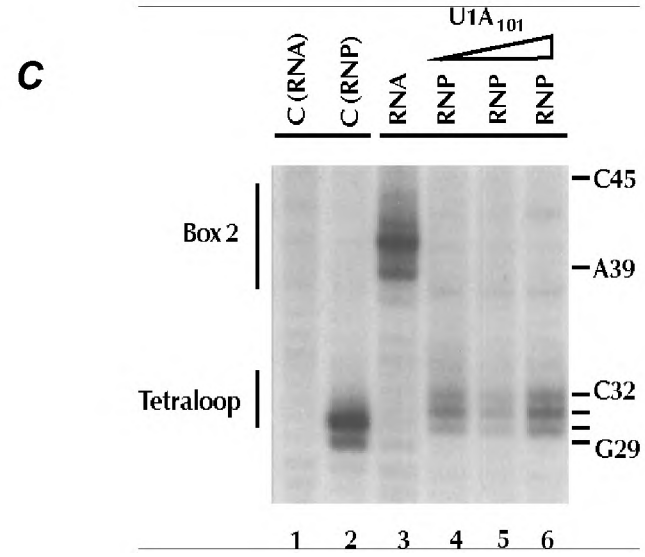


Figure 4 (cont.) *(C) Enzymatic footprinting of the $U1A_{101}$ - Ag RNA complex using T2. The region of the tetraloop is shown. Lanes 1 and 2 are control lanes (incubations of 1: only RNA; 2: RNA and protein) in which the enzyme was omitted . Lane 3 shows the enzymatic probing of the naked RNA, while lanes 4-6 show probing of the complex by increasing amounts of $U1A_{101}$ (50, 100, and 150 times excess, respectively)*

cleaves C31 and C32 which are both not cleaved in the naked RNA (Figure 4C, lane 3). Note that this effect seems to be more prominent for $U1A_{101}$ than for U1A (Figure 4C, lanes 4-6 and Figure 4A, lanes 3 and 4).

The 3' part of the structure (stem 4 and loop 4) does not show much protection (see Figures 4A and 4C), so this region appears to be largely accessible in the RNP complex. Only the 5' side of the polyadenylation signal (nts. A64 and A65) shows partial protection (Figure 4A, lanes 2 and 4).

The footprinting experiments using Fe(II)EDTA yielded no distinct protection pattern. Both the free RNA and the U1A bound RNA showed equal reactivity towards the hydroxyl radicals at all positions (data not shown). Fe(II)EDTA probing can be used to detect three-dimensional folding of the RNA molecule if riboses are in the interior of the RNA molecule and therefore protected from strand-scission.[148] Our results of the naked RNA showed equal reactivity for all nucleotides, indicating that this RNA has no tertiary interactions.

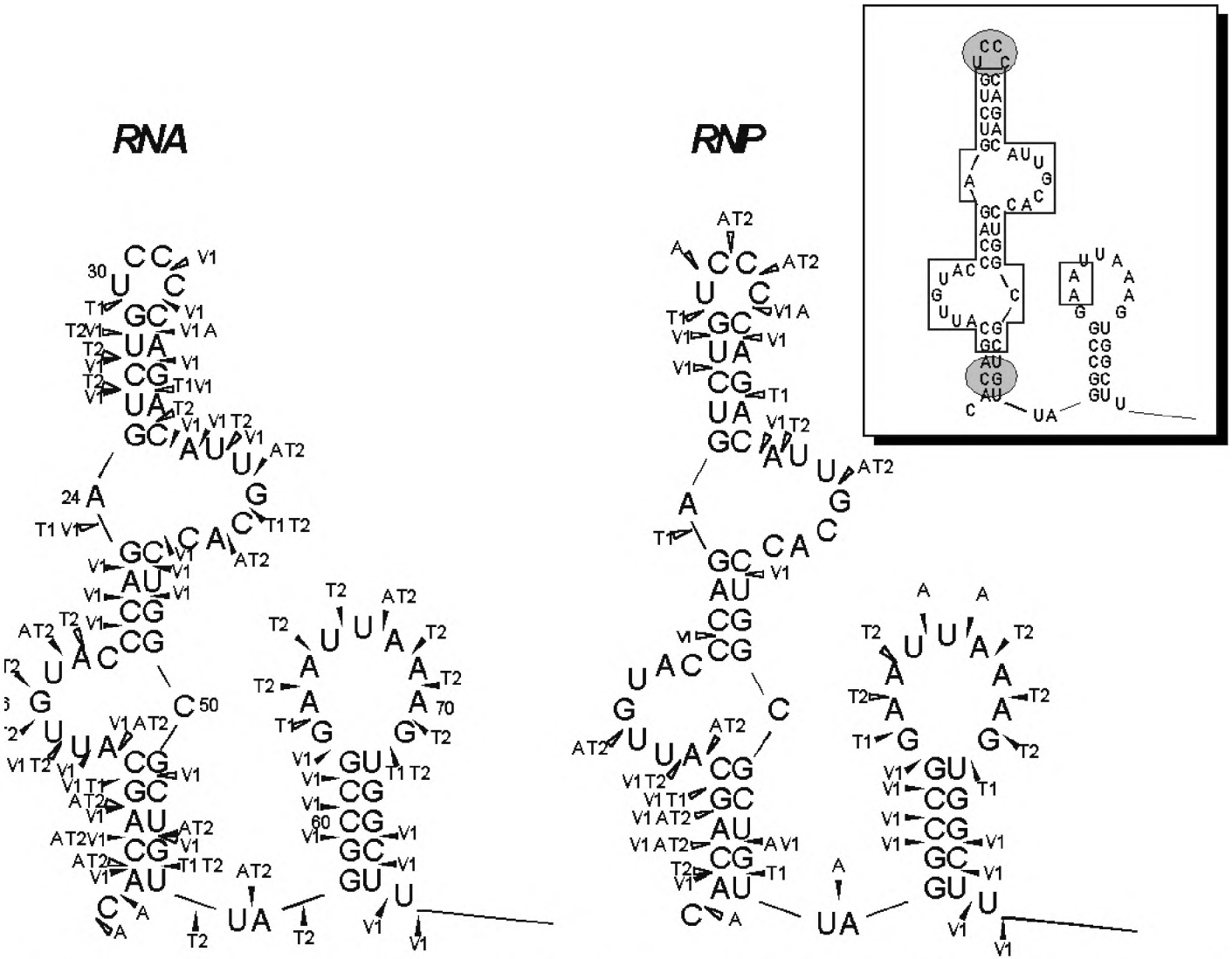
DISCUSSION

Secondary structure of Ag RNA

The conserved region of the 3' UTR of Ag RNA was probed at nucleotide resolution by the utilization of chemical probes. The secondary structure thus obtained further refines the structure predicted previously, which was based upon enzymatic probing and analysis of structure and function of mutant RNAs.[121] At 20 °C, the highly conserved stems 2 and 3 are indeed present while stems 1 and 4 also seem to exist, although not very stable and probably breathing. At 0 °C, all four stems clearly exist.

At 20 °C all nucleotides in the Box 1 and 2 regions are fully accessible at both their Watson-Crick positions and at the N7-atoms of the purines. This behavior excludes the presence of tertiary interactions between these nucleotides and other parts of the RNA. At 0 °C, several nucleotides in the Box regions are no longer accessible, probably because of stacking. Interestingly, the behavior of the two Box regions at 0 °C is not identical. Box 1 shows more reactivity of both Watson-Crick and N7-positions at the 5' end of the Box, while Box 2 shows most reactivity at the 3' end of the Box region. A (somewhat) different structure of the two Boxes could be expected because the two sequences, although almost identical in sequence, have a different structural context in the Ag RNA and also differ in U1A binding affinity. Box 2 forms a much stronger (~30 fold) binding site for U1A protein than Box 1.[121]

The two unpaired nucleotides A24 and C50 are clearly accessible at 20 °C. Whether the accessibility of A24 and C50 results from looping out of the helix or from the fact that the structure of the Ag RNA is more open at the internal loops, cannot be deduced from these results. The fact that N7 of G25 can also be modified suggests that A24 is not stacked in the helix.



ont.). Summary of Rnase footprinting data. *(D) Summary of RNase data obtained for both the naked Ag RNA A/U1A$_{101}$-Ag RNA complex. On the left the digestion pattern of the RNA is shown (adapted from van Gelder et al., hile on the right the digestion pattern of the U1A/U1A$_{101}$-Ag RNA complex is shown. Strong cleavages are indicated ows and weak cleavages by open arrows. The enzymes which cut are indicated next to the arrows. The insert rotected areas indicated by boxes and the regions of structural changes indicated by shaded circles.*

Cleavage in the tetraloop by RNases was not observed,[121] but chemical probing showed that in the RNA three of the four nucleotides are moderately accessible at 20 °C for the much smaller chemical probes. N3 of C31 was only reactive under denaturing conditions, which could either indicate stacking or an involvement in tertiary interactions under native conditions. At 0 °C, none of the nucleotides in the tetraloop are accessible.

The chemical probing results clearly indicate the presence of stem-loop 4. In the loop containing the AUUAAA polyadenylation signal, it can be seen that all adenosines are reactive at both the N1 and the N7 position, both at 20 °C and at 0 °C.

In conclusion, the probing studies provide a secondary structure for the Ag RNA as shown in Figure 1. The data also suggest that there are no major tertiary interactions within the naked RNA. Almost all purines are accessible at their N7-position. This is supported by our Fe(II)EDTA experiments which show equal reactivity of all nucleotides in the Ag RNA, indicating the absence of tertiary interactions. A similar behavior towards the Fe(II)EDTA generated hydroxyl radicals was found in the 5S rRNA.[148]

The Ag RNA-U1A protein complex

Footprinting experiments have been performed on the complex of Ag RNA and the U1A protein using several ribonucleases and Fe(II)EDTA. Inhibition of reactivity at certain nucleotides can be inferred as a direct protection (and hence contact) of the RNA by the protein at that site. However, reduced reactivity can also be caused by conformational changes in the RNA chain brought about by the addition of the protein, and it is difficult to distinguish between these two modes of reduced reactivity. Furthermore, since RNases are large molecules, steric hindrance may significantly enlarge the protected regions. Footprinting experiments were also conducted with $U1A_{101}$, which contains only the N-terminal RNP motif of U1A, but these experiments yielded the same results.

The protection experiments show that all 7 nucleotides contained in the Box 1 and Box 2 regions, as might be expected, are largely protected when the U1A protein is bound. Clearly, these sequences, which have been shown to be important for U1A binding to U1A mRNA (15), are in contact with the U1A protein. Only some nucleotides, located at the 5'-side of both Boxes, can be attacked by the ribonucleases.

All nucleotides in stems 1, 2, and 3 show complete or partial protection against ribonucleases in the presence of U1A protein, and also in the presence of $U1A_{101}$. Around nucleotide 54, RNase V1 cleavage is enhanced when U1A protein is added. This indicates that stem 1 becomes more stable as a result of U1A binding. Alternatively it could be that stacking of stems 1 and 4 on each other occurs. This is supported by the absence of RNase T2 cleavages for nucleotides G53 - G58. The reduction of cleavages by single-strand-specific RNases in stem 1 and 3 could indicate protection of these regions by the U1A protein, but could also be the result of a further stabilization of the stems upon binding of U1A protein. Footprinting of the complex of stem-loop II of U1 snRNA and U1A protein has been performed with both RNase V1

and ethylnitrosourea.[106] In that case only one of the stem halves of stem-loop II appeared to be protected against the probes. This difference in behavior of U1 snRNA [106] as compared to Ag RNA can be due to the difference in size of the RNA substrates or to the presence of two rather bulky U1A proteins in the latter case, instead of one in the case of U1 snRNA. The environment of the recognition sequence also differs between the two RNA substrates. In the case of U1 snRNA the seven nucleotides are flanked by three loop nucleotides with a flexible conformation, whose only function is to allow the seven nucleotide recognition sequence to adopt the conformation necessary for protein binding.[149] In the Ag RNA the seven nucleotides are flanked by stem structures and a bulged nucleotide. Obviously, these nucleotides are less flexible and therefore might induce a different structure of the RNA, leading to a different protection pattern.

Interestingly, in the RNP complex the nucleotides in the tetraloop (U30-C33) seem to become accessible. This may indicate that this loop undergoes a considerable conformational change upon U1A protein binding. It appears that the loop opens up, with its nucleotides becoming available to the probes. Although we cannot completely exclude that the bands appear as a result of secondary enzyme cuts in the RNA molecule, this seems not very likely. The observed protection pattern in Box 2 indicates that the U1A protein is still able to bind the Ag RNA (Figure 4C, compare lanes 3 and 4-6), suggesting that the structure of the complex is still intact. If the bands in the tetraloop are secondary enzyme cuts, the complex would probably be disrupted. Similar behavior of a tetraloop is also found in bacteriophage R17 where the tetraloop structure is altered upon R17 coat protein binding.[150] Our results are seemingly in contrast with NMR data recently published,[124] in which it was shown that the tetraloop structure remained unperturbed upon protein binding. However, in those experiments a modified Ag RNA was used which contained only Box 2 flanked by two stem structures and hence, able to bind only one U1A protein. More importantly, stem 3 was shortened by one base pair and the tetraloop had a different sequence as in the original Ag RNA used by us, in order to stabilize stem 3. The recently described mechanism for complex formation does not describe a possible conformational change in the tetraloop [118] but in this study the same modified RNA was used as in the study mentioned before.[124]

The 3' part of the structure is formed by stem-loop 4 and this part shows, as expected, no protection, except for some limited changes at the 5' side of the loop. This means that this region is largely accessible in the RNP complex. This is in agreement with the finding that U1A does not interfere with the binding of the cleavage polyadenylation specificity factor (CPSF) to the polyadenylation signal during polyadenylation of the U1A pre-mRNA.[122]

The crystal structure of the U1A - U1 snRNA complex [89] revealed that the recognition surface of the protein binds the RNA in a “stem-downwards” manner.[88] Since the recognition sequence on the Ag RNA is the same, the most likely binding mode of the U1A protein to this RNA will be similar. A theoretical model has been proposed for the complex of two binding domains of U1A and the Ag RNA (excluding

stem-loop 4) using the crystallographic data combined with molecular modeling.[125] Our experimental data show that stem 2 is protected, both by the U1Awt as well as by the $U1A_{101}$ protein. This indicates that the N-terminal RNP motif is sufficient for protecting stem 2 and possibly involved in the protein-protein interactions. Our experimental data are in complete agreement with the theoretical model which predicts protein-protein interactions between Lys96 and Asp24, as well as between Lys60 and Gln39.[125]

The model for the U1A-3'UTR complex could also explain the results of the Fe(II)EDTA experiments. Since the protein-protein interactions occur via the major groove of stem 2 this leaves the minor groove exposed to the solvent. Furthermore, the minor grooves of stem 1 and stem 3 are exposed to the solvent. Fe(II)EDTA generated hydroxyl radicals attack the riboses of the RNA in the minor groove, which therefore would not result in any protection pattern.

In conclusion, we have obtained detailed experimental information concerning the structure of Ag RNA and its complex with U1A protein. These data can aid in elucidating the process of autoregulation of the U1A protein.

ACKNOWLEDGEMENTS

The authors want to thank Drs. Pascale Romby and Chantal Ehresmann for advice with the chemical probing reactions and Dr. Ger Pruijn for critical reading of this manuscript.

MATERIALS AND METHODS

In vitro transcription of RNA and purification of recombinant U1A protein - *In vitro* transcription by T7 RNA polymerase was carried out as described.[121] The conserved region of the 3' UTR of the U1A mRNA is called Ag RNA and was cloned into the EcoRI and HindIII sites of pGEM-3Zf(+) resulting in Ag RNA transcripts. The nucleotide sequence of the Ag fragment of U1A extends from VI-842 to VI-951 in the sequence [143] and includes 8 nucleotides at the 5' end derived from the vector plasmid. Production of recombinant U1A protein and of the mutant protein $U1A_{101}$ (consisting of the first 101 N-terminal amino acids of U1A protein) in *E. coli* was carried out as described.[120]

5'- and 3'-end-labelling - For 5'-end-labelling the dephosphorylated RNAs or oligodeoxynucleotides were labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase as described previously.[121] For 3'-end-labelling, the RNAs were labeled using [^{32}P]pCp and T4 RNA ligase as described.[136]

Chemical Modification - Concentrations of chemicals were optimized to obtain single hit conditions. Control incubations, in which the reagent was omitted, were always performed in parallel to detect spontaneous pyrimidine-purine breaks, which easily occur in RNA,[52,144] and, in the case of the primer extension method, also to detect spontaneous stops of reverse transcriptase. Chemical modifications were performed both on unlabelled (0.3-0.5 µg) and on 3'-end-labelled RNA (3×10^4 cpm), which was always renatured before use. The RNAs were modified under native conditions (N; presence of magnesium), semi-denaturing conditions (SD; presence of EDTA) and denaturing conditions (D; high temperature, presence of EDTA). Native and semi-denaturing reactions were conducted both at 20 °C and at 0 °C. Modification reactions with DMS (dimethylsulfate) and CMCT (1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-p-toluene sulfonate) were essentially carried out as described.[136] It should be noted that DMS and CMCT as well as the chemicals mentioned below are hazardous and should be handled with proper precautions. Chemically modified

nucleotides were detected either by primer extension analysis or by using 3'-end-labelled RNA (in the case of N7-G, N7-A and N3-C).

In case of probing the N7-G positions with DMS, an additional aniline treatment was carried out to produce strand scission at the site of modification.[52]

During the DEPC (diethylpyrocarbonate) treatment, 10-60 µl DEPC was added to the sample in 200 µl Buffer N or SD (final concentration 0.3 - 1.9 $x10^{-9}$ M; incubation for 1 hr at 20 °C). For D conditions, 3-10 µl DEPC was added to 200 µl Buffer SD and incubated for 7 min at 90 °C (final concentration 0.09 - 0.3 $x10^{-9}$ M). Reactions were stopped by ethanol precipitation. An aniline step was performed to produce strand scission at the site of the modification.[52]

Primer extension analysis - Primer extension was carried out as described.[136] Oligodeoxynucleotide primers used were 5'-GCTTAACAGCGCCAGG-3' and 5'-GATTGTGAAAAACCAAACCTC-3', complementary to nucleotides 45-60 and 81-101 in Ag RNA, respectively. Reverse transcripts were analyzed on 10% denaturing polyacrylamide gels.

Enzymatic footprinting - After renaturation of the 5'-end-labelled Ag RNA (3 x 10^4 cpm, final concentration ~6 nM) the specified amount of U1A wt protein (150 - 300 fold excess) or $U1A_{101}$ (50 - 300 fold excess) was added (final volume: 20 µl). Buffer conditions were 10 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$ and 50 mM KCl. The complex was allowed to form for 30 min at 20 °C after which the probing reactions were performed with RNase T1 (0.15 U; U1A wt only), RNase A ($1x10^{-5}$ U; U1A wt only), RNase T2 (0.005 U), or RNase V1 (0.06 U) as described.[121]

Fe(II)EDTA footprinting - The reactions were essentially carried out as described,[87] with varying concentrations of Fe^{2+}, EDTA, ascorbic acid and hydrogen peroxide.

3

Characterisation of an anti-RNA recombinant autoantibody fragment (scFv) isolated from a phage display library and detailed analysis of its binding site on U1 snRNA

This is the first study in which the complex of a monoclonal autoantibody fragment and its target, stem-loop II of U1 snRNA, was investigated with enzymatic and chemical probing. A phage display antibody library derived from bone marrow cells of an SLE patient was used for selection of scFvs specific for stem-loop II. The scFv specificity was tested by RNA immunoprecipitation and nitrocellulose filter binding competition experiments. Immunofluorescence data and immunoprecipitation of U1 snRNPs containing U1A protein, pointed to a scFv binding site different from the U1A binding site. The scFv binding site on stem-loop II was determined by footprinting experiments using RNase A, RNase V1 and hydroxyl radicals. The results show that the binding site covers three sequence elements on the RNA, one on the 5' strand of the stem and two on the 3' strand. Hypersensitivity of three loop nucleotides suggests a conformational change of the RNA upon antibody binding. A three-dimensional representation of stem-loop II reveals a juxtapositioning of the three protected regions on one side of the helix, spanning approximately one helical turn. The location of the scFv binding site on stem-loop II is in full agreement with the finding that both the U1A protein and the scFv are able to bind stem-loop II simultaneously. As a consequence, this recombinant monoclonal anti-U1 snRNA scFv might be very useful in studies on U1 snRNPs and its involvement in cellular processes like splicing.

S. W. M. Teunissen, M. H. W. Stassen,
G. J. M. Pruijn, W. J. van Venrooij, R. M. A. Hoet.
*RNA, **4**, 1124-1133 (1998)*

Introduction

Patients suffering from autoimmune diseases produce antibodies directed to a variety of intracellular self-antigens. In patients with systemic lupus erythematosus (SLE), antibodies recognising the U1 small nuclear ribonucleoprotein particle (snRNP) are often found. U1 snRNP consists of an RNA molecule (U1 snRNA) bound by several proteins and plays an important role in splicing of pre-mRNA. Three proteins are specifically associated with U1 snRNP: U1A, U1C and the U1-70K protein. The remaining U1 snRNP proteins (Sm proteins) are also associated with other snRNPs (for a review see Klein Gunnewiek et al., 1997).[151]

The sera of SLE patients may not only contain antibodies directed against the protein components of the U1 snRNP, but also antibodies directed against the U1 snRNA itself.[152] Previous studies have shown that the main targets of these autoantibodies are stem-loops II and IV of U1 snRNA. Hoet et al. (1992) [153] utilised mutant RNAs to study these epitopes on the U1 snRNA, while Tsai & Keene (1993) [154] and St. Clair & Burch (1996) [155] employed an RNA epitope library. Both anti-U1 snRNA antibody populations (anti-stem-loop II and anti-stem-loop IV) have been purified and were able to precipitate the native U1 snRNP, indicating that the antigenic regions are accessible.[156] Stem-loop II and in particular the loop of this structural element is involved in the binding of the U1A protein.[104-109] A protein interacting with stem-loop IV has not yet been described.

Both human and murine monoclonal antibodies directed to U1 snRNP protein components are available for studying the U1 snRNP particle.[157-159] Monoclonal antibodies against the U1 snRNA are not available yet. Thus far, only affinity purified antibodies or *in situ* hybridisation of modified anti-sense probes could be used to study the U1 snRNP via its RNA molecule.[156,160] Therefore, we set out to produce an anti-U1 snRNA monoclonal antibody. For this purpose a phage display antibody library was prepared from an anti-U1 snRNA positive SLE patient and subsequently used to select anti-U1 snRNA antibodies. These selections resulted in the isolation of two single chain variable fragments (scFv) specific for stem-loop II of U1 snRNA. Like the patients' antibodies, the selected scFvs were able to bind intact U1 snRNPs. One of the scFvs was further characterised in detail. The scFv binding region on U1 snRNA was determined by footprinting analyses using RNase A, RNase V1 and Fe(II)EDTA. Finally, the scFv binding site was visualised in a three-dimensional model of stem-loop II RNA.

Results

Selection of U1 snRNA specific scFvs. A phage display antibody library was prepared from SLE patient Z5 whose serum contained antibodies directed against U1 snRNA (Figure 1A), more specifically anti-stem-loop II and anti-stem-loop IV antibodies.[156] First, RNA was isolated from bone marrow lymphocytes. Subsequently, immunoglobulin cDNA was synthesised and amplified using RT-PCR to obtain a scFv

phage display library.[161] From this library two scFvs against U1 snRNA, termed Z5scFv3 en Z5scFv7, were selected using procedures described previously.[159] Soluble scFvs were produced as a periplasmic secretion product containing both a VSV-tag and a His-tag for purification and detection purposes.

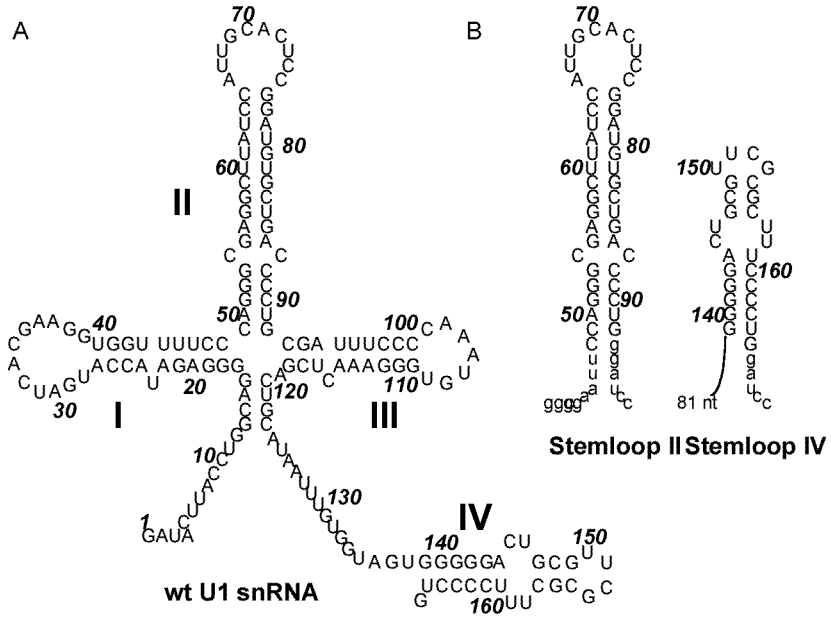


Figure 1.*(A) Human U1 snRNA in its proposed secondary structure according to Krol et al., (1990).[163] Stem-loop structures I to IV are indicated.(B) Stem-loop subfragments of U1 snRNA. Stem-loop II (left) and stem-loop IV (right) are shown. Vector derived nucleotides are indicated in lower case. The numbering of the nucleotides is according to U1 snRNA.*

The specificity of the scFvs was first established using RNA immunoprecipitation. Since previous studies had revealed that patient Z5 produced anti-U1 snRNA autoantibodies reactive with stem-loop II and stem-loop IV of U1 snRNA, the corresponding subfragments of U1 snRNA (see Figure 1B) were used to determine the specificity of the isolated scFvs. ScFvs were incubated with a mixture of [^{32}P]-labelled RNAs, consisting of U1 snRNA, stem-loop II and stem-loop IV subfragments, and Th RNA (the RNA component of RNase MRP [162]) as a negative control. Both Z5scFv3 and Z5scFv7 were able to immunoprecipitate U1 snRNA and stem-loop II, but not stem-loop IV (Figure 1C). The negative control, Th RNA, was not precipitated. These results show that both scFvs specifically recognise stem-loop II of U1 snRNA. Nitrocellulose filter binding competition assays, using several RNAs and double stranded DNA, confirmed the specificity of these scFvs for stem-loop II (data not shown). Possible crossreactivity of the scFvs was tested with U1 proteins U1A, U1C, U1-70K, and SmB/B' (by Western blotting and ELISAs) and with nuclear extracts of HeLa cells (by Western blotting). No crossreactivities were detected (data not shown).

Since both scFvs were reactive with stem-loop II and the expression level of Z5scFv7 was (for unknown reasons) reproducibly low, Z5scFv3 was chosen for a more detailed characterisation.

Characterisation of Z5scFv3. To determine which part of stem-loop II is recognised, the binding of Z5scFv3 to two mutants of stem-loop II was analysed by immunoprecipitation. In mutant stem-loop II^{ML} four loop nucleotides, which are part of the U1A binding sequence (65-AUUG-69), were replaced by UGAU (stem-loop II^{ML}, mutated loop; Figure 1D).

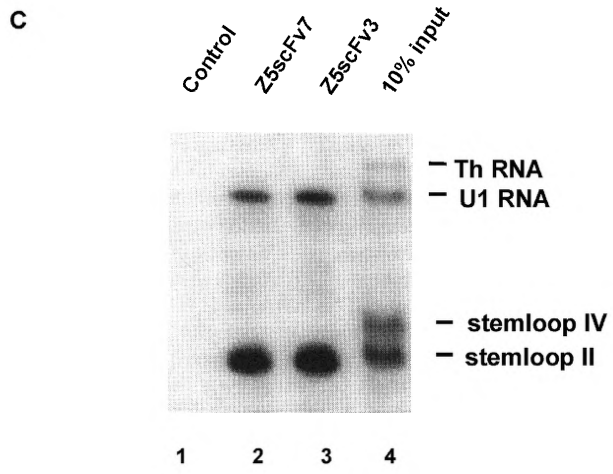


Figure 1 (Cont.). *(C) Immunoprecipitation of U1 snRNA, stem-loop II, stem-loop IV, and Th RNA. The stem-loop II subfragment used in this experiment contains 45 extra vector derived nucleotides, resulting in an RNA molecule of 89 nucleotides. Lane 1 is a control incubation with anti-VSV tag antibodies coupled to protein A beads via rabbit anti-mouse antibodies. Lane 2 and 3 show RNAs precipitated by Z5scFv7 and Z5scFv3, respectively. Lane 4 shows 10 % of the RNA input.*

As can be seen in Figure 1D (lane 3), Z5scFv3 was able to precipitate stem-loop II^{ML}. In the second mutant the stem of stem-loop II is replaced by stem IV of U2 snRNA (stem-loop II^{MS}, mutated stem; Figure 1E). In this case Z5scFv3 failed to precipitate the RNA (Figure 1E, lane 3). Taken together, these results indicate that Z5scFv3 specifically binds to the stem structure of stem-loop II of U1 snRNA.

Next was tested whether the Z5scFv3 was able to recognise its epitope in intact U1 snRNPs. First, a co-immunoprecipitation experiment was performed. The scFv was incubated with a ^{35}S-methionine-labelled cell extract and the co-immunoprecipitated proteins were analysed by SDS-PAGE followed by autoradiography. As a positive control anti-Sm antibodies were used (Figure 1F, lane 2). Both types of antibodies were able to precipitate the U1 snRNP complex, visualised by the presence of the U1A and U1C proteins as well as the SmB/B' and SmD proteins (see Figure 1F, lanes 2 and 3). In a second type of experiment, the immunofluorescent staining pattern for the scFv in HEp-2 cells was determined. As can be seen in Figure 2, Z5scFv3 resulted in either a coiled body staining pattern or in a combined coiled body/nucleoplasmic speckled

pattern very similar to that obtained with anti-U1A monoclonal antibody 9A9. Also Z5scFv7 resulted in staining of the coiled bodies. The appearance of nucleoplasmic speckles in addition to coiled bodies appeared to be dependent on the concentration of the antibody fragment. Similar observations have been made before with other anti-snRNP antibodies and with anti-U1C scFvs.[161] The staining of both nucleoplasm and coiled bodies was sensitive to RNase activity (results not shown). Since Z5scFv3 is specific for stem-loop II of U1 snRNA, this result again suggests that the region of stem-loop II that is recognised by the scFv is different from the region that is bound by the U1A protein. From these experiments we conclude that both U1A and the Z5scFv3 autoantibody fragment are able to bind simultaneously to stem-loop II of U1 snRNA.

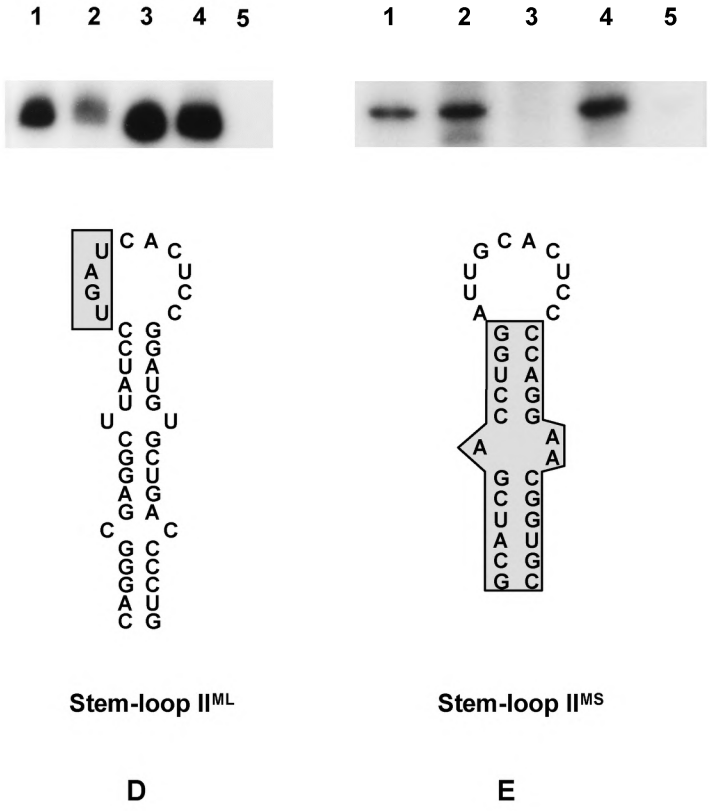


Figure 1 (Cont.). *(D) Immunoprecipitation of mutant stem-loop II RNA. The mutant stem-loop $II^{ML\ 153}$ contain a mutated U1A binding sequence, replacing 65-AUUG-69 with UGAU (depicted by a shaded box). Lane 1 shows the input. Lanes 2 and 3 show the supernatant and the precipitated RNA by Z5scFv3, respectively. Lanes 4 and 5 show the supernatant after immunoprecipitation and the precipitated RNA of the control incubation with anti-VSV tag antibodies coupled to protein A beads via rabbit anti-mouse antibodies.(E) Immunoprecipitation of mutant stem-loop II RNA. In this mutant stem II is changed into the stem of stem-loop IV of U2 RNA (depicted by a shaded box), creating mutant stem-loop II^{MS} .[153] Lane 1 shows the input. Lanes 2 and 3 show the supernatant and the precipitated RNA by Z5scFv3, respectively. Lanes 4 and 5 show the supernatant after immunoprecipitation and the precipitated RNA of the control incubation with anti-VSV tag antibodies coupled to protein A beads via rabbit anti-mouse antibodies.*

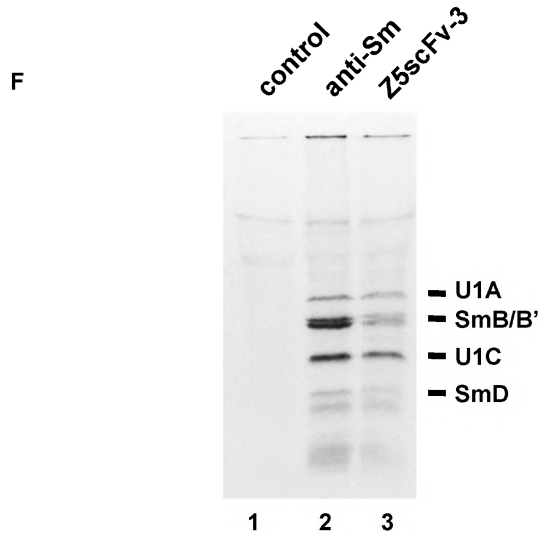


***Figure 1 (Cont.)**. (F) Precipitation of intact U1 snRNPs by Z5scFv3. U1 snRNP particles containing ^{35}S-labeled proteins were immunoprecipitated from a whole cell extract using Z5scFv3 (lane 3), anti-Sm antibodies (lane 2) and a non-related antibody as negative control (lane 1).*

Probing of the U1 snRNA - Z5scFv3 complex - For a more detailed characterisation of the binding site of the patient derived scFv, stem-loop II of U1 snRNA (Figure 1B) was probed with both enzymatic and chemical reagents in the presence and the absence of Z5scFv3. The scFv was first bound to beads and allowed to bind to stem-loop II. The bound stem-loop II was subsequently probed with RNase A, RNase V1 or Fe(II)EDTA generated hydroxyl radicals.

Enzymatic footprinting - Three regions on the RNA were found inaccessible to RNase V1 (cleaves phosphodiester bonds in double stranded RNA) upon binding of the scFv, as can be seen in the experiment shown in Figure 3A (compare lanes 2-4 with lanes 6-8). The first region extends from G53 to U61 (nucleotide numbering as in Figure 1B), covering a large central portion of the stem. The other regions, G77-G82 and G92-G93, are located on the opposite strand of the RNA (Figure 3A), although it should be noted that the interpretation for G93 was obscured by spontaneous RNA breaks in some experiments. These results, however, confirm that the binding site of the scFv is different from the U1A binding site. No data were obtained for the region G83 to G91, since RNase V1 did not cleave the free nor the complexed RNA at these positions. The probing results for the free RNA molecule are in complete agreement with the RNase V1 probing data for U1 snRNA described by Krol et al. (1990).[163] The data are summarised in Figure 4A.

An unexpected observation was the increased sensitivity to RNase V1 of the bonds 3' of C73-C75 in the RNA-scFv complex (see Figure 3A, lanes 3 and 4, and Figure 4A). This suggests that binding of the scFv induces stacking of the bases of the last three loop nucleotides, making them susceptible to RNase V1 cleavage. This could be

the result of a more general stabilisation of the helix structure, suggesting that the binding of Z5scFv3 to stem-loop II induces some structural changes in the RNA.

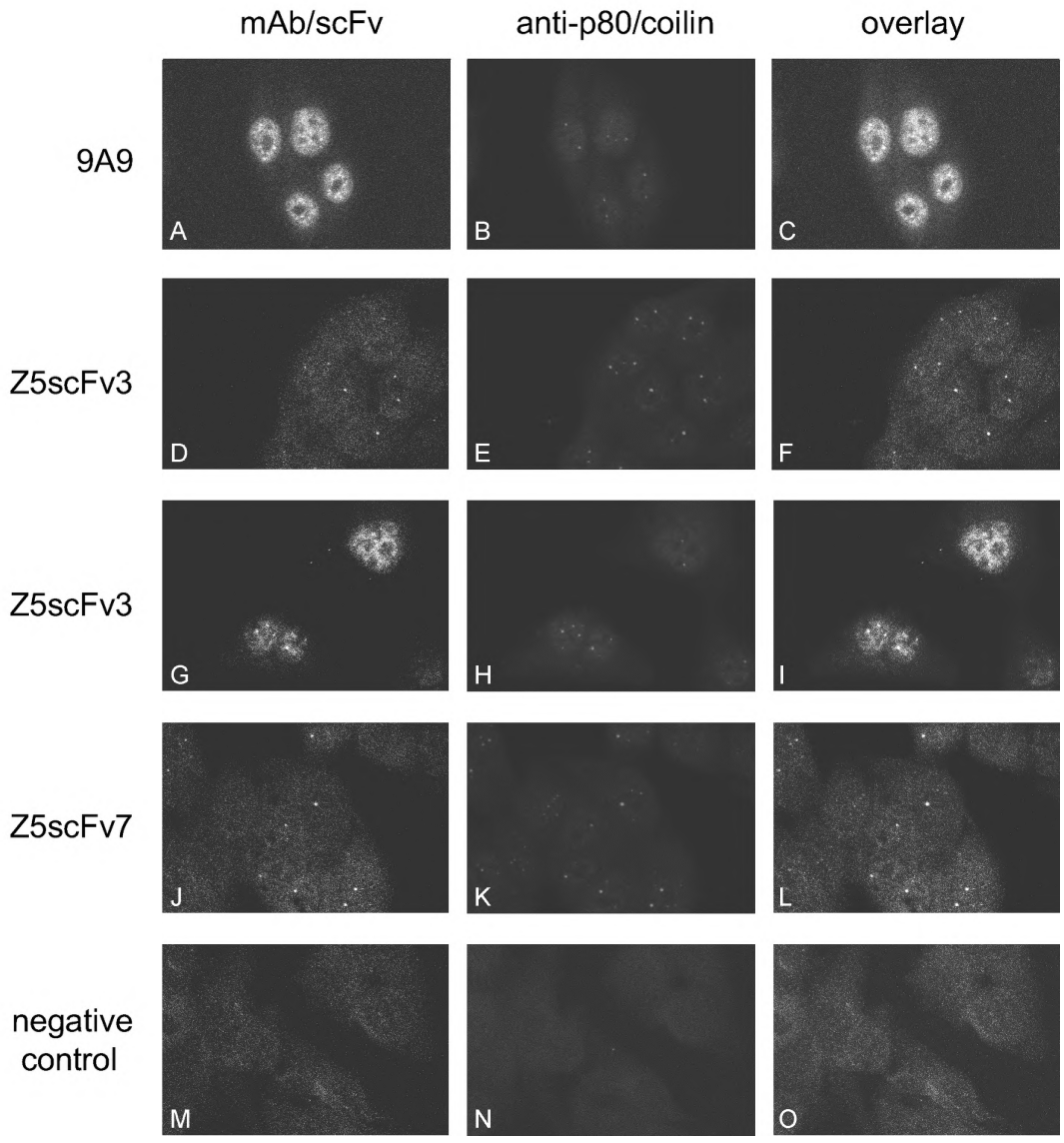


Figure 2. The anti-stem-loop II antibodies stain coiled bodies and nucleoplasmic speckles. *Immunofluorescent staining of HEp-2 cells with 9A9 (A, as a positive control), Z5scFv3 (D and G), Z5scFv7 (J) all showing green fluorescent patterns (denoted mAb/scFv), and anti-p80/coilin (B, E, H, and K, for co-localisation studies) showing red fluorescence. Panels C, F, I, L, and O show the overlay of the two different antibody stainings, in which co-localisation results in a yellow colour. HEp-2 cells were incubated with antibody (in G-I a five-fold higher concentration of Z5scFv3 was used, compared to D-F) and stained subsequently with anti-VSV-G (Boehringer Mannheim) and TexasRed conjugated rabbit-anti mouse Ig (Dako). As a negative control, the incubation with mAb/scFv and anti-p80/coilin was omitted (M-O). A colour version will be available via www.kun.nl*

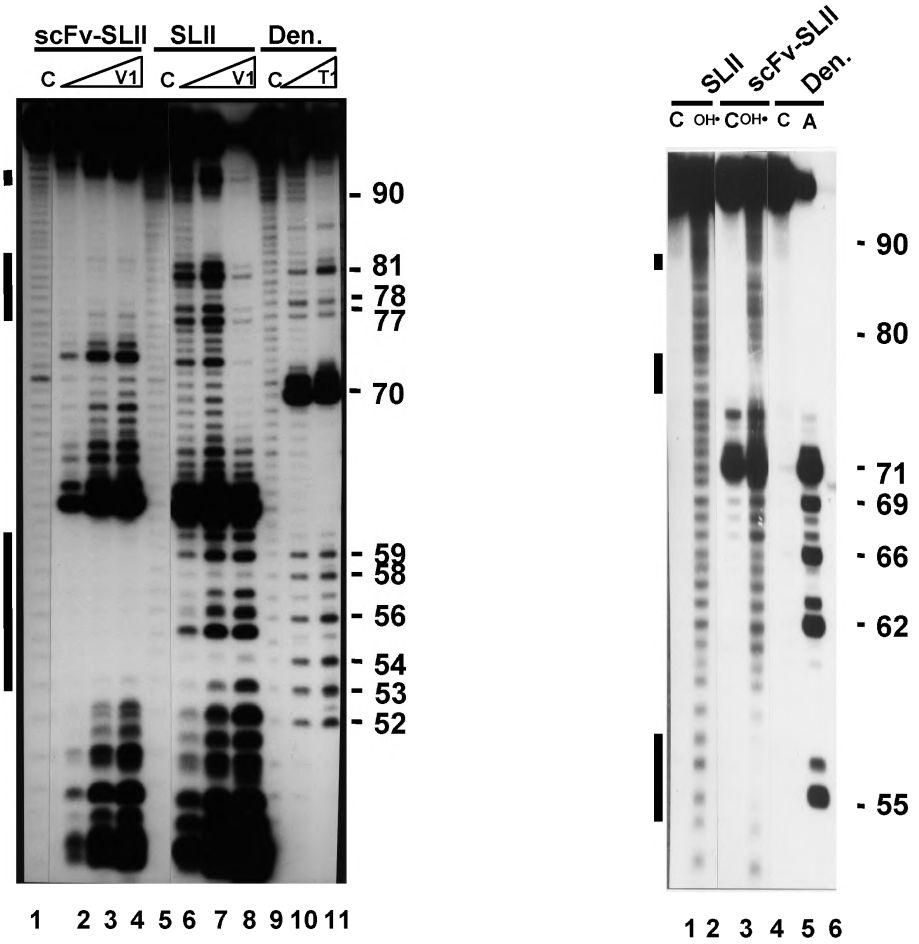


Figure 3 Footprint analysis of the Z5scFv3-stem-loop II complex*.(A) Probing experiment using the double-stranded RNA specific RNase V1. Both the free and the complexed RNA were subjected to RNase V1 degradation. Lanes 1, 5 and 9 are control incubations in which the RNase was omitted. The Z5scFv3-stem-loop II complex is probed in lanes 2-4 (denoted scFv-SLII), with increasing amounts of RNase V1. The free RNA is probed with increasing amounts of RNase V1 in lanes 6-8 (denoted SLII). Cleavage positions were deduced from sequence ladders of RNase T1 (lanes 10 and 11). It should be noted that the cleavage product of RNase V1 migrates faster than the corresponding RNase T1 cleavage product, resulting in a difference of one nucleotide. The nucleotide positions indicated on the right are numbered according to the RNase V1 cleavages. The solid lines at the left indicate regions in the RNA which are protected by Z5scFv3.(B) Probing experiment using Fe(II)EDTA. Lanes 1 and 2 show probing of the free RNA (denoted SLII), lanes 3 and 4 show probing of the Z5scFv3-stem-loop II complex (denoted scFv-SLII), and lanes 5 and 6 show the sequence ladder of the free RNA obtained under denaturing conditions using RNase A. Lanes 1, 3, and 5 are control incubations in which the reagent was omitted. Cleavage positions were deduced from the sequence ladder of RNase A (lane 6). It should be noted that the labelled product from cleavage by hydroxyl radicals has an additional phosphate group. The migration therefore differs by one nucleotide from the RNase A sequence ladder. The nucleotide positions indicated on the right are numbered according to the hydroxyl radical cleavages. The solid lines at the left indicate regions in the RNA which are protected by Z5scFv3.*

Probing with RNase A (cleaves 3' of single stranded pyrimidines) did not show a protection pattern. The enzyme was able to cleave the same bonds in both the RNA-scFv complex and the free RNA, i.e. those in the loop of the RNA (nt 66-74), which is in

agreement with the data described above suggesting that the scFv binds to the stem region on stem-loop II (data not shown).

Chemical footprinting - Since ribonucleases are relatively large molecules, the protected regions might appear larger due to steric hindrance. We therefore used hydroxyl radicals generated by Fe(II)EDTA to obtain more detailed information on the accessibility of the riboses. An example of such a Fe(II)EDTA probing experiment is shown in Figure 3B. Lanes 1 and 3 represent control incubations in the absence of Fe(II)EDTA. A comparison of lanes 2 and 4 reveals three regions in stem-loop II that are protected by the bound scFv against hydroxyl radical induced cleavage. The combined phosphorimaging data of Fe(II)EDTA experiments were used to calculate a percentage protection of each nucleotides (upper panel of Figure 4B). These data delineate the binding region of Z5scFv3 found by RNase V1 to nucleotides C54-G58 on the 5' strand of the stem and to G77-U80 and C87-C88 on the 3' strand of the stem (lower panel of Figure 4B).

If the Z5scFv3 binding site is highlighted in the three-dimensional model of the stem,[163] all protected nucleotides appear to reside on the same side of the stem. The binding site encompasses three juxtaposed sequence elements on the helix (nts. 54-58, nts. 77-80, and nts. 87-88). This is shown as stereoviews from two different angles (Figure 5A and 5B), with the Z5scFv3 binding site highlighted in red. The loop II structure of U1 snRNA, as determined by X-ray crystallography,[89] is incorporated in the model (shown in yellow) together with the first RNA binding domain of U1A protein (represented as a green shaded ribbon). This model shows that the Z5scFv3 binding site is spatially separated from the U1A binding site and is located at the other side of the RNA molecule (with respect to the helix axis). This model further suggests that both the U1A protein and Z5scFv3 are able to bind the RNA simultaneously.

Discussion & Conclusion

In this study human monoclonal autoantibodies directed to an RNA molecule have been described. Their specificity for U1 snRNA, in particular stem-loop II of U1 snRNA, was established by immunoprecipitation and by nitrocellulose binding competition assays. One of these human monoclonal antibodies (Z5scFv3) was characterised in more detail. It specifically binds to the stem of the U1 snRNA stem-loop II, suggesting that the scFv recognises a structure distinct from the U1A binding site, which is known to bind the loop of stem-loop II.[104-109] The simultaneous binding of Z5scFv3 and U1A to U1 snRNA was substantiated by immunofluorescence data and by the ability of the autoantibody fragment to precipitate U1A containing U1 snRNPs from a total cell extract. The present data, however, do not provide information about the possible difference in affinity of the scFv towards either the intact U1 snRNP or the naked U1 snRNA. Nevertheless, the binding of U1 snRNP in both immunoprecipitation and immunofluorescence imply that this scFv might be an attractive tool for biochemical and cell biological studies on U1 snRNP. In view of its specificity it provides an alternative to U1 snRNA-specific anti-sense oligonucleotide probes.[160] For

some applications it may even be superior to anti-sense oligonucleotides, because the latter probes are directed preferentially to single-stranded regions in the RNA, which in general coincide with functionally important elements, like the 5'-end of U1 snRNA, which is involved in 5'-splice-site recognition, and loop sequences, which are important for protein binding (reviewed in Klein Gunnewiek et al., 1997 [151]). Since no (major) structural changes in the RNA seem to be required for recognition of U1 snRNA by the scFv and since no specialised and laborious techniques are required to visualise U1 snRNA with the scFv, this recombinant antibody fragment provides an excellent reagent for cell biological studies on U1 snRNA. Moreover, since in principle any epitope tag can be introduced in the scFv (a VSV-G tag was used in our studies), these antibodies can be easily combined with conventional murine monoclonal antibodies in double staining experiments.

It should be noted that we still do not know whether the Z5scFv3 epitope is accessible in higher order complexes like the spliceosome. Nevertheless, the immunofluorescence patterns obtained with the anti-U1 RNA scFvs indicated that the stem-loop II epitope is accessible in both nucleoplasmic U1 RNA and U1 RNA localised in coiled bodies. At relatively low concentrations of scFvs merely coiled body staining was evident, which is in agreement with previously published data indicating that the highest concentrations of splicing snRNPs are found in this subnuclear compartment, an organelle which has been proposed to be involved in some aspect of snRNP maturation, transport, or recycling.[164] The anti-U1 RNA scFv staining pattern is similar to those reported previously for anti-U1 snRNP specific antibodies [158,165] and to the pattern obtained by *in situ* hybridisation with U1 snRNA-specific probes. Most interestingly, based upon the results of the latter type of experiments, it has been suggested that the relative concentration of U1 snRNA in coiled bodies might be lower than that of the other splicing snRNAs.[160,166,167] The present scFv data suggest that also U1 snRNA accumulates in the coiled bodies in HEp-2 cells. The discrepancy with the *in situ* hybridisation data might be explained (in part) by the potentially reduced accessibility of U1 snRNA sequences complementary to the *in situ* hybridisation probes in coiled bodies.

The RNase V1 structure probing data of U1 snRNA stem-loop II are in complete agreement with previously published results obtained with complete U1 snRNA.[163] These data and the immunoprecipitation of stem-loop II indicate that the vector derived nucleotides did not (markedly) influence the formation of the proposed stem-loop structure. Our RNase V1 probing data do suggest that the stem structure is lengthened somewhat by base pairing of some of these vector derived nucleotides (see Figure 4A). A comparison of the RNase V1 cleavages in the free RNA with cleavages in the scFv-RNA complex, indicated three protected regions in the complex. One region is located in the 5' part of the stem of stem-loop II, and two regions in the 3' part. A reduced efficiency of a particular cleavage is not necessarily the result of steric hindrance due to the binding of the scFv. It could also result from a structural change, i.e. from double-stranded to single-stranded RNA. However, the RNase A probing of the scFv-RNA complex did not show additional cleavage sites in regions that are

double-stranded in the free RNA, thus arguing against such a possibility. Because of the size of nucleases, footprints found by RNases usually are larger than the region actually involved in the interaction with protein.

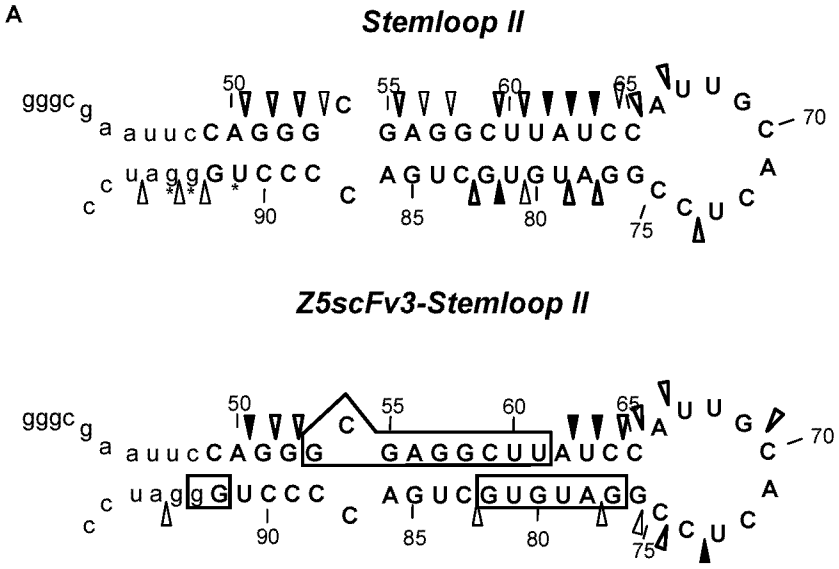


Figure 4 Summary of footprinting data. *(A) Summary of probing data of RNase V1 cleavages. Enzymatic cleavages are indicated by open triangles (weak), open bold triangles (normal) and solid triangles (strong). RNA breaks which occurred in the control incubations are indicated with asterisks. Protected regions are indicated by boxes. Vector derived nucleotides are indicated in lower case.*

A more refined protection pattern, nicely correlating with the RNase V1 data, was obtained with hydroxyl radicals (Figure 4B). It should be noted that the Z5scFv3 binding region contains some unusual base pairs: C54-C87, G55-A86, and A56-G85. Since such base pairs distort the normal helical (A-form) structure of an RNA, they might be very important for the specific recognition of the RNA by the scFv. An interesting possibility would be that the major groove, which is deep and narrow in an A-form helix and therefore inaccessible for protein side chains, is widened to some extent by the presence of these unusual base pairs. Specificity is more easily achieved by interactions in the major groove in comparison with minor groove contacts, in particular in the case of A-U base pairs. Since the centre of the epitope recognised by Z5scFv3 coincides with the unusual base pairs (positions 54-58), it is tempting to speculate that the specificity of the antibody fragment is determined by hydrogen bonds and specific electrostatic interactions in this unique (distorted) helical region in stem-loop II of U1 snRNA. Structural data, obtained by for instance x-ray diffraction of co-crystals, are required to shed more light on this.

To obtain a more three-dimensional impression of the Z5scFv3 epitope structure, a model for U1 snRNA stem-loop II was built, based upon literature data for stem [163] and loop structure.[89] A projection of sequence elements protected by Z5scFv3 on this three-dimensional representation of U1 snRNA stem-loop II revealed that these

elements are juxtaposed on one side of the helix and that the complete epitope covers approximately one helical turn. Strikingly, in this structural model the U1A protein recognition site, which has been well defined by many experimental data,[89,104-109] is not only spatially separated from the Z5scFv3 epitope through its location in the loop, but also appears to be positioned on the opposite side (with respect to the helix axis) of the RNA molecule. Accordingly, it is not surprising that U1A and Z5scFv3 are able to bind to stem-loop II simultaneously.

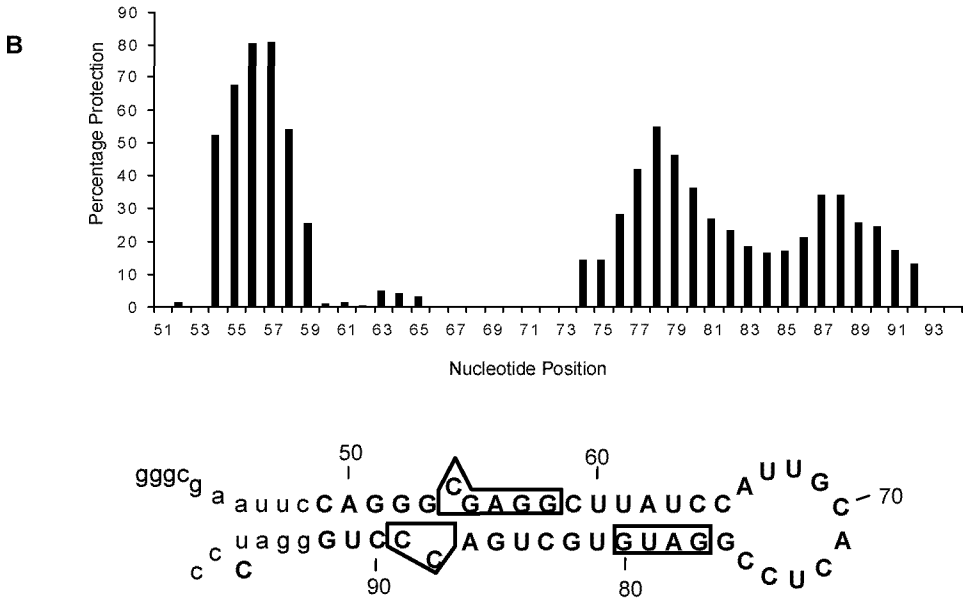


Figure 4 (Cont.). *(B) Summary of Fe(II)EDTA probing data. The upper panel shows the average protection for each nucleotide, calculated from phosphorimaging data. Protected regions above 30 % protection are boxed in the lower panel. It should be noted that the data for nucleotides 67-75 were obscured by natural RNA breaks in the control incubations.*

An interesting aspect from an immunological point of view is the fact that Z5scFv3 was derived from a phage display library reflecting the antibody repertoire of an autoimmune patient. The observation that this autoantibody recognises part of the stem structure of stem-loop II, even in a completely assembled U1snRNP complex, is compatible with the idea that the intact U1snRNP complex, at some stage during the development of the autoimmune response, is the (auto)antigen exposed to the immune system.

Acknowledgements

Use of the services and facilities of the Dutch National NWO/SURF Expertise Centre CAOS/CAMM, University of Nijmegen, The Netherlands, is gratefully acknowledged. The co-ordinates of the three-dimensional model of U1 snRNA were generously supplied by Dr. E. Westhof (I.M.B.C., CNRS, Strasbourg, France). The

authors wish to thank Y. van Aarssen and M. Pieffers for technical assistance and Dr. E. Chan (Scripps Research Institute, La Jolla, USA) for the anti-p80 coilin antiserum.

These investigations were supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Technology Foundation (STW grant 349-3026).

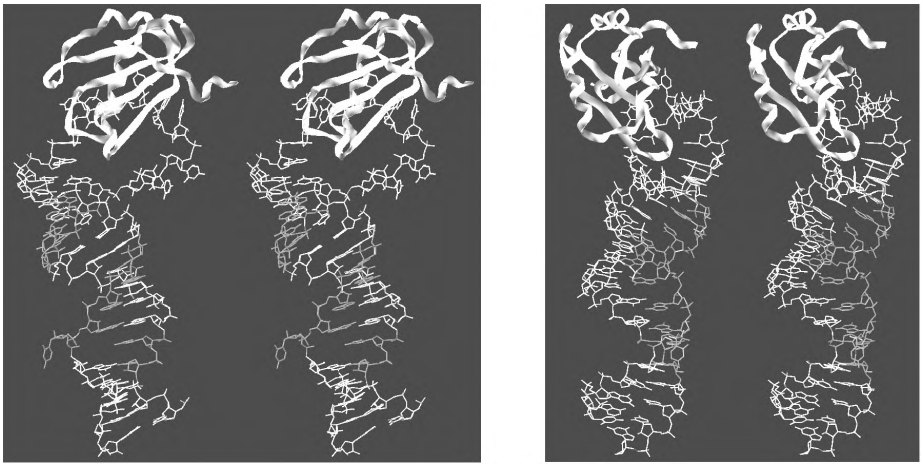


Figure 5. Stereoview of the binding site of Z5scFv3 on stem-loop II. *Front view (A) and side view (B) of stem-loop II (reconstructed from the crystal structure of the U1A binding site[89] and the proposed three-dimensional model of stem-loop II [163]). The U1A binding sequence on the RNA is indicated in yellow. The protected regions on stem-loop II are shown in red. The RNA binding domain of the U1A protein in association with the loop is shown as a green shaded ribbon. A colour version will be available via www.kun.nl.*

Materials and Methods

Preparation of RNA and production of the U1A protein - The full length U1 snRNA, the stem-loop II and the stem-loop IV subfragments (Figure 1) are described by Hoet et al. (1992).[153] The constructs for stem-loop II and IV code for nucleotides C49-G92 and G139-G165, respectively. After T7 transcription of the BamHI digest, stem-loop II contains 15 extra vector derived nucleotides (see Figure 1B) and 45 extra vector derived nucleotides for the subfragment used in Figure 1C, while stem-loop IV contains 86 extra vector derived nucleotides (see Figure 1B). The Th RNA construct was a kind gift of Dr. H. Pluk. Transcription and 5'-end-labelling of the RNA was performed essentially as described in Teunissen et al. (1997).[168] To obtain ^{32}P-labelled RNA, *in vitro* transcription was performed in the presence of [α-^{32}P]UTP. Production and purification of recombinant U1A protein in *E. coli* was carried out as described previously.[120]

Selection and screening of scFvs - To obtain a scFv producing library, RNA was isolated from bone marrow lymphocytes. Subsequently, cDNA was synthesised and amplified using PCR.[89] The selection protocol followed was essentially as described in De Wildt et al. (1996),[159] using immunotubes (Nunc, Maxisorp) coated with U1 snRNA to select scFvs. After three rounds of selection, scFv-clones were screened by a nitrocellulose filter binding assay.

^{35}S-labelled cell extract - HeLa (ATCC-CCL-2) monolayer cells were cultured in methionine-free medium for one hour. Subsequently, the cells were incubated with 10 μCi/ml ^{35}S-methionine for four hours at 37°C, followed by an overnight incubation in complete medium. Cells were trypsinized

and washed three times with cold PBS. After resuspension, the cells were subjected to three rounds of freeze-thawing, sonified and centrifuged at 100.000 g. Glycerol was added to the supernatant to a final percentage of 20% which was subsequently stored at -70°C until use.

Immunoprecipitations - A VSV tagged scFv was coupled to 20 µl protein A agarose beads via rabbit anti-mouse immunoglobulins (Dako, 5 µl) and a mouse anti-VSV antibody (Boehringer Mannheim, 1667351; 0,5 ml supernatant) in 150 mM NaCl, 10 mM Tris-HCl, pH 7.5 and 0.05% NP-40.[159] In order to test the specificity of the antibodies, the immobilised scFvs were either incubated with a mixture of [^{32}P]-labelled RNAs (Th RNA, U1 snRNA, stem-loop II and stem-loop IV) or with the [^{32}P]-labelled stem-loop II^{ML} or stem-loop II^{MS}. To determine whether Z5scFv3 was able to immunoprecipitate U1 snRNP complexes, the coupled antibody was incubated with ^{35}S-labelled cell extract. For probing reactions, the immobilised Z5scFv3 was incubated with stem-loop II RNA at 4°C in probing buffer: 100 mM KCl, 10 mM Tris-HCl pH 7.5, 2.5 mM $MgCl_2$, 0.1% Tween-20 and 0.1% NP-40. The complexes were washed with probing buffer to remove unbound RNA and subsequently subjected to probing agents.

Immunofluorescence staining of HEp-2 cells - Fixed HEp-2 (ATCC-CCL-23) cells were incubated with a periplasmic fraction of Z5scFv3 or Z5scFv7 (dialysed against PBS) and bound antibody fragments were visualised by subsequent incubations with anti-VSV-G (Boehringer Mannheim) and FITC-conjugated goat anti-mouse Ig (Dako). Between each incubation step the cells were briefly washed three times with PBS. The scFv was omitted in the negative control. The U1A specific monoclonal antibody 9A9 [158] was used as a positive control. For double-labelling experiments with anti-p80 coilin (rabbit serum R288, a kind gift of Dr. E. Chan) the cells were incubated with anti-p80 coilin (100-fold diluted in PBS) and Texas Red-conjugated donkey anti-rabbit Ig (Amersham) either prior to the scFv incubation or in combination with the scFv and the FITC-conjugated goat anti-mouse Ig, respectively. Confocal microscopy was used to obtain Z scans, which were superimposed for double stainings.

Probing reactions - Concentrations of enzymes and chemicals were optimised to obtain single hit conditions. Parallel incubations in which the probing agent was omitted were performed in order to detect spontaneous RNA breaks. RNase probing was performed essentially as described by Teunissen et al. (1997).[168] Probing with Fe(II)EDTA was carried out as described by Darsillo & Huber (1991).[87] In order to obtain single hit conditions, the concentrations of $Fe(II)SO_4$ (Aldrich), EDTA (Aldrich), ascorbic acid (Sigma) and H_2O_2 (Janssen Chimica) were varied, while maintaining a molar ratio of 1:2 for $Fe(II)SO_4$ and EDTA, respectively. The results were analysed on a phosphorimager. The average percentage protection was calculated by subtracting the number of counts of the Z5scFv3-stem-loop II complex from the number of counts of the naked RNA, after background correction, for each nucleotide position.

Construction of the three-dimensional RNA structural model - Structures were visualised on a Silicon Graphics Indy workstation with the SYBYL program (SYBYL Molecular Modeling Software, Version 6.3/6.4, Tripos Inc., St.Louis, Mo., USA). The model was constructed using PDB structure 1NRC for the loop [89] and the three-dimensional model proposed by Krol et al. (1990) [163] for the stem structure. The base pairs of the stem from the crystal structure were manually fitted onto the corresponding base pairs of the model.

4

Conserved features of Y RNAs: a comparison of experimentally derived secondary structures

In this study, phylogenetically conserved structural features of the Ro RNP associated Y RNAs were investigated. The human, iguana, and frog Y3 and Y4 RNA sequences have been determined previously and the respective RNAs were subjected to enzymatic and chemical probing to obtain structural information. For all of the analysed RNAs, the probing data were used to compose secondary structures, which partly deviate from previously predicted structures. Our results confirm the existence of two stem structures, which are also found at similar positions in hY1 and hY5 RNA. For the remaining parts of hY3 and hY4 RNA the secondary structures differ from those previously proposed based upon computer predictions. What might be more important, is that certain parts of the RNAs appear to be flexible, i.e. to adopt several conformations. Another striking feature is that a characteristic pyrimidine-rich region, present in every Y RNA known, is single-stranded in all secondary structures. This may suggest that this region is readily available for base pairing interactions with other cellular nucleic acids, which might be important for the as yet unknown function of the RNAs.

S. W. M. Teunissen, Martijn J. M. Kruithof, A. Darise Farris,
John B. Harley, Walther J. van Venrooij, Ger J. M. Pruijn
Submitted for publication.

Introduction

Ro ribonucleoprotein particles (Ro RNPs) are evolutionarily conserved RNAs which are associated with the Ro60 and La proteins.[133] In humans, the RNA component of a Ro RNP can be one of four different Y RNAs: hY1, hY3, hY4, and hY5. Although the existence of Ro RNPs was discovered almost two decades ago and the Y RNAs as well as the Ro60 and La proteins seem to be well conserved during evolution, no function for these RNP complexes has as yet been found.[133] Recently, however, a synergistic role for the Ro60 and La proteins in regulating the translation of L4 ribosomal protein mRNA has been suggested in *Xenopus laevis*.[126] Most interestingly, a yet unidentified RNA was found to be associated with this complex, suggesting that the whole Ro RNP could be involved.[126]

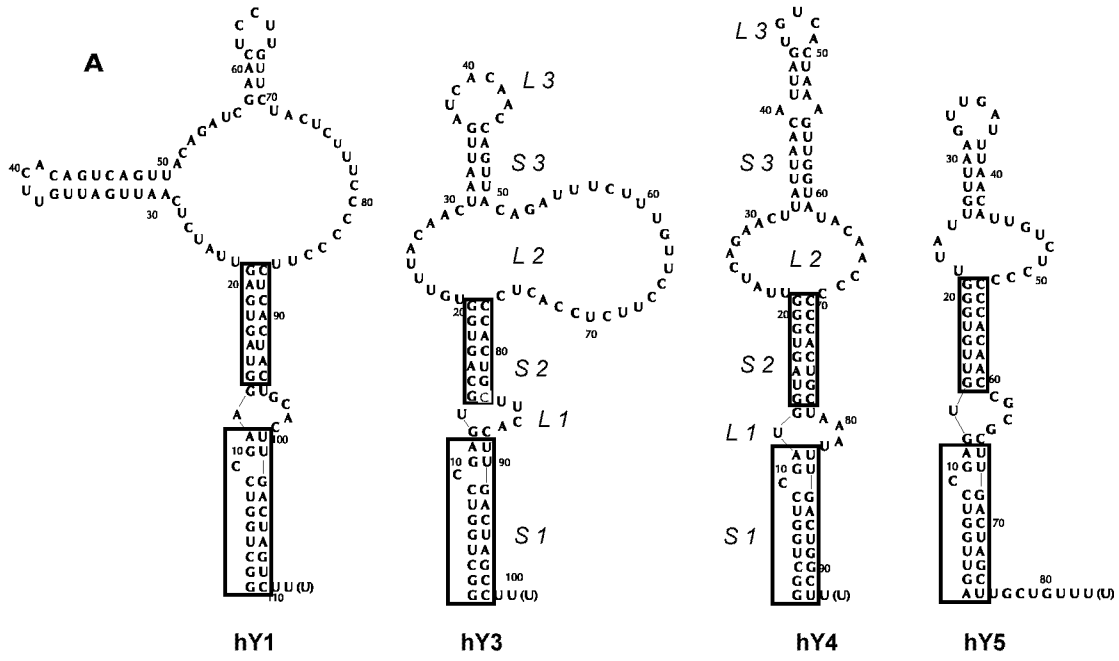


Figure 1. Secondary structure of the human Y RNAs and sequence alignment of Y3 and Y4 sequences. *(A) Proposed secondary structure of the four different human Y RNAs, adapted from.136 The conserved lower stem structures are boxed. The stem loop structures are indicated (stem 1: S 1; loop 2 L 2; etc.).*

Several Y RNA sequences from a variety of organisms have been described.[133-135,137,169-172] The small RNA molecules are predicted to fold into a conserved secondary structure containing at least three stem structures (Figure 1A). A small internal loop separates stem 1 and stem 2, while a larger pyrimidine-rich internal loop is positioned between stem 2 and stem 3.[173] Stem 1 contains the most highly conserved sequences, corresponding to the Ro60 binding site.[128,130] Sequence variations in Y RNAs of different organisms occur mostly in the large internal loop and the third stem structure.[133,173]

The structures of hY1 and hY5 RNA have been studied most extensively by both enzymatic and chemical probing experiments.[136] Hardly any experimental data are available on the structures of hY3 and hY4 RNA, although, as is the case for the hY1 and hY5 RNAs, phylogenetic data,[135-137] as well as a recently described algorithm,[173] support the predicted secondary structures. In order to obtain more knowledge on their structures and possibly more insight in the conservation of structural elements, we set out to determine experimentally the (secondary) structures of Y3 and Y4 RNA from human, frog, and iguana via enzymatic and chemical probing experiments. The probing results will be compared with computer predictions and combined with phylogenetic data.

Results

The secondary structure predictions and the available phylogenetic and experimental data for the Y RNAs revealed a number of conserved structural elements. Figure 1A shows the proposed secondary structures for the four human RNAs, which are based upon a combination of those data. A general feature of these structures are two highly conserved stem structures (Figure 1A, stems 1 and 2), which are separated by a small internal loop 1. Loop 2, varying in size from 13 to 36 nucleotides, contains a characteristic pyrimidine stretch (most prominent in Y1 and Y3 RNA). A third stem structure (stem 3) is associated with this internal loop in all Y RNAs.[173]

In contrast to hY1 and hY5 RNA,[136] these structures for hY3 and hY4 RNA are only marginally supported by experimental probing data. Therefore, the latter structures were addressed in this study and we extended these structure probing experiments to Y3 and Y4 from two additional organisms, namely frog and iguana.

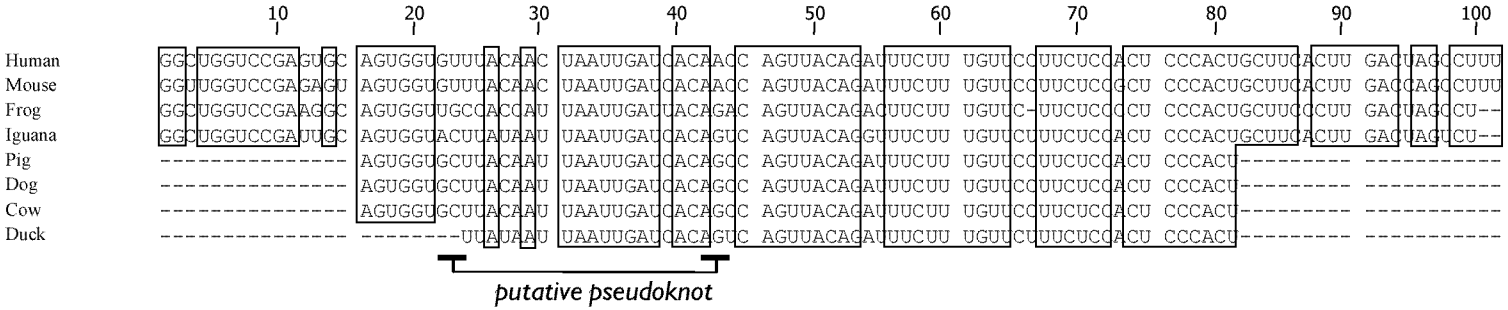
Y3 RNA

The (partial) sequences of Y3 RNA homologues have been determined from eight different organisms so far. An alignment of these sequences (Figure 1B) shows large clusters of conserved residues (indicated by boxes). The sequences are listed according to the degree of sequence similarity with the human sequence. The mouse sequence only differs by five nucleotides from the human sequence and was therefore not analysed by chemical or enzymatic probing.

Figure 1 (Cont., opposite page). (B) *Sequence alignment of the known Y3 RNA sequences. The conserved regions are boxed. The nucleotide numbering is according to the human sequence. For the pig, dog, cow, and duck sequences, the dashed lines indicate the primer sequence used in the PCR reaction to isolate the Y RNA Note that only four of these sequences are full length. The flanking sequences are missing in the remaining four sequences (indicated by dashes), because of the use of oligo primers to detect these sequences {169}. Hence, no sequence data for those regions is available yet. Below the alignment, a possibility for a pseudoknot is indicated.* ***(C)*** *Sequence alignment of the known Y4 RNA sequences. The conserved regions are boxed. The nucleotide numbering is according to the human sequence. For the pig, dog, and monkey sequences, the dashed lines in front and after the sequence (1-16 and 76-93), indicate the primer sequence used in the PCR reaction to isolate the Y RNA {169}. Hence, no sequence data for those regions is available yet.*

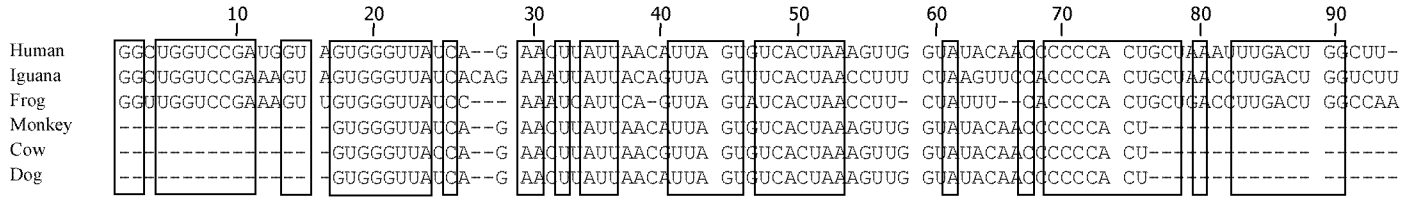
B

Y3 RNA sequence alignment



C

Y4 RNA sequence alignment



hY3 probing - First, *in vitro* transcribed, 5'-end radiolabelled hY3 RNA was subjected to enzymatic digestion under native conditions with RNase V1 (specific for dsRNA), RNase T2 (ssRNA), RNase T1 (ssG), and RNase A (ssU & ssC). During probing experiments, parallel incubations without RNases were used to detect any spontaneous RNA cleavage products. RNase T1 and/or RNase A digestions under denaturing conditions were included to pinpoint nucleotide positions.

Figure 2A shows the results for probing hY3 RNA with RNase A and RNase V1 under different concentrations of Mg^{2+}. Lanes 6 through 10 show probing under native conditions (100 mM KCl, 2.5 mM $MgCl_2$) at 20°C. The summarized probing data are shown in Figure 3A. Of the stem structures, stem 1 and stem 2 are supported by the enzymatic probing data. Stem 2, as can be seen in Figure 2A, lanes 9 and 10, shows almost exclusively RNase V1 cleavages. However, stem 3 (nts. 30-50) is cleaved by both RNase A and RNase V1. The existence of a stable stem 3 under native conditions is therefore not evident.

The pyrimidine-rich internal loop (loop 2, see Figure 1A), is of particular interest in view of its potentially functional relevance. As was previously found by Van Gelder and co-workers [136] for the corresponding region in hY1 RNA, the loop 2 region in hY3 shows only a few, relatively weak RNase A cleavages. A possible explanation for the reduced RNase A sensitivity might be that these nucleotides are involved in the formation of a stable and compact tertiary interaction. Furthermore, hY1 probing experiments revealed that the accessibility of loop 2 was highly dependent on the Mg^{2+} concentration (Teunissen, unpublished observations). Probing experiments with hY3 RNA using different concentrations of Mg^{2+} demonstrated a similar behavior for the corresponding region of hY3 RNA (Figure 2A). The simultaneous increased intensity of RNase V1 cleavages and decreased intensity of RNase A cleavages at higher Mg^{2+} concentrations in this region of the RNA (nts. 61-74) indeed suggest the formation of a Mg^{2+} dependent tertiary interaction. Also, C45 appeared to be less accessible at higher magnesium concentrations. Unexpectedly, RNase V1 cleavages in stem 1 (nts. 1-12) were reduced at higher Mg^{2+} concentrations (data not shown). Together, these data indicate significant magnesium dependent structural changes in loop 2 of hY3 RNA.

As described, the region of stem 3 shows both cleavages ssRNA and dsRNA specific enzymes, which complicates the interpretation. There are a number of possible explanations for the ambiguous behavior of this part of hY3. One possible explanation for both single-strand and double-strand specific enzyme cleavages might be that the molecules are folded into a mixture of different conformations in equilibrium. This could lead to single-strand specific cleavages from one conformation and, for the same region, double-strand specific cleavages from another conformation. Although both breathing of unstable stems and folding into different conformations might occur simultaneously the DMS probing results and the migration of the RNA as a single species in native polyacrylamide gels (unpublished results) suggest that breathing is the main cause for structural heterogeneity in the central part of hY3 RNA. A second explanation could be that it contains double-stranded regions which are not very stable and thus 'breathing' under the experimental conditions. Such a dynamic structural

heterogeneity would provide a suitable target for both kinds of enzymes. Breathing of stem structures is a phenomenon that is reduced if the temperature is lowered. Therefore, we probed hY3 RNA using DMS, both at 0°C and at 20°C, the results of which are shown in Figure 2B. Residues modified by DMS were detected by reverse

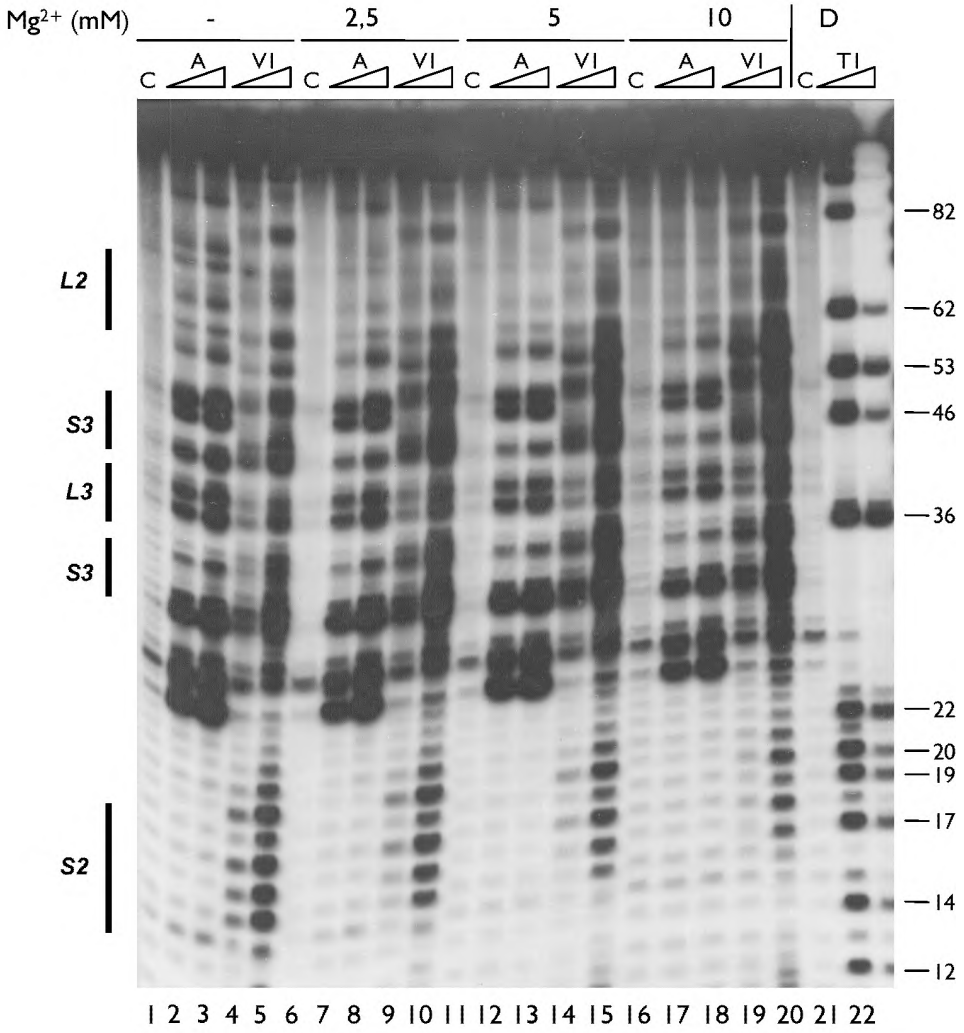


Figure 2. Secondary structure probing experiments on Y3 RNA. *(A) Probing experiment on hY3 RNA, using RNase V1 and RNase A under different concentrations of Mg2+. Lanes 1-5 show probing without Mg2+, lanes 2 & 3 using RNase A, lanes 4 & 5 using RNase V1, and lane 1 is a control incubation in which the nuclease was omitted. Similarly, lanes 6-10 show probing under 2,5 mM Mg2+, lanes 11-15 under 5 mM Mg2+, and lanes 16-20 under 10 mM Mg2+ (indicated above the autoradiograph). Lanes 21 and 22 show a denaturing control incubation and an incubation with RNase T1, which are used to pinpoint nucleotide positions (indicated on the right). Note that the RNase V1 cleavage products lack the 3' phosphate group and therefore migrate slightly slower though the gel. The position of loop 2 and stem 2 is shown on the left.*

transcription using a primer complementary to the most 3' region of the RNA. As frequently observed for this mode of detection, also in this case there are many spontaneous stops of the reverse transcriptase reaction, both after chemical modification at 0°C and at 20°C (Figure 2B, lanes 1 and 2, respectively), which

obscured the data for a number of nucleotides. Nevertheless, the results clearly show that at 20°C, also many modifications by DMS occurred in predicted double-stranded regions (see Figure 3A for a summary). The probing results at 0°C generally show a diminished intensity of bands corresponding to modifications, indicating a possible general stabilisation of base-paired structures noted that the reactivity of DMS at 0°C will also be lower. However, some nucleotides are relatively less accessible at 0°C. For example, in comparison with nucleotide A40 nucleotide A46 is less intensely modified at 0°C. This indicates that A40 is single-stranded and that A46 is located in a double-stranded structure which is not very stable at 20°C, but which is stabilised at lower temperatures.

Since the probing data were not completely consistent with the proposed secondary structure, we decided to calculate a number of alternative secondary structures using MFOLD.[174] The unambiguous probing data were used to discriminate between the nine calculated structures and resulted in the structure shown in Figure 3A. The structure of the conserved stems 1 and 2 including loop 1 is slightly different from the previously proposed structure (compare Figure 1A and Figure 3A). Loop 1 consists of nts. A87-U90, which was predicted in seven of the nine structures generated by MFOLD using the parameter settings described in Materials and Methods. The existence of stems 1 and 2 is supported by the RNase V1 data. A major difference with the previously proposed structure occurs in loop 2 and stem loop 3 where, in the new structure, larger regions are predicted to be double-stranded. Nucleotides A28-U31 form a stem structure with A46-U49, which is supported by the RNase V1 cleavages at C27-A29 and at C45-G47. It should be noted that RNase V1 requires a region of at least two nucleotides on either side of the hydrolysis site adopting an approximately helical conformation.[54] Although the RNase V1 data support the existence of this stem structure, it consists of three A-U base pairs and one G-C. The stem is therefore probably not very stable and breathing or is affected by different conformations. This is indeed supported by the single-strand specific enzyme cleavages seen at nucleotides A46-U49. Similarly, the stem formed by nucleotides U34-G36 and C41-A43 shows both single-strand and double-strand specific cleavages, probably due to a breathing stem or structural heterogeneity. The latter two stems are separated by an internal loop of four nucleotides, which might have an additional negative effect on the stability of these two stems. Another stem is formed by A52-A54 and U68-U70, which is again supported by the RNase V1 data, but not by single-stranded specific cleavages. The structure of the remaining loop is still unclear, since both types of enzymes cleave in similar regions. The structure shown in Figure 3A might be one of several conformations. An interesting observation in this respect is that hY3 might be able to form a pseudoknot structure between G22-U24 and C44-A42. RNase V1 data support the pseudoknot structure. The existence of base pairing between these residues is further supported by phylogenetic data, as can be seen in Figure 1B. The mouse sequence is identical to the human sequence in this region. The other sequences, with the exception of frog, have complementary mutations that still allow, and thus support, formation of a pseudoknot. Formation of this pseudoknot

might induce a large conformational change in the molecule. The most important result presented here, is that except for stems 1 and 2, the secondary structure is probably very dynamic.

xY3 probing – We subjected xY3 RNA to enzymatic probing experiments using RNase V1 and RNase T2 to globally asses its secondary structure. An example of probing with RNase V1 is shown in Figure 2C. The cleavages indicate a large number of double-stranded regions, spanning from G17 to C73. Also these data, together with RNase T2 probing data, are summarised in Figure 3B, left panel.

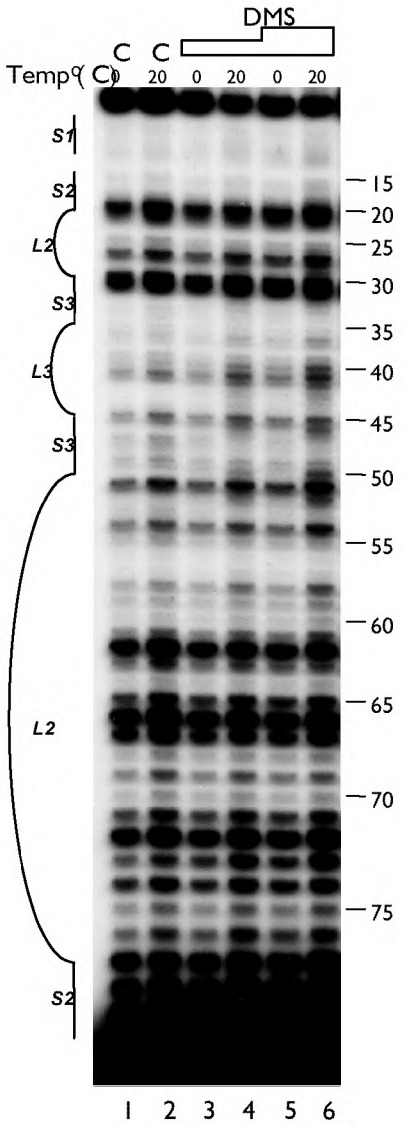


Figure 2 (Cont.). *(B) Probing hY3 RNA with DMS at 0°C (lanes 1, 3, and 5) and 20°C (lanes 2, 4, and 6). Sites of modification were detected using primer extension. Parallel control incubations without DMS were used to detect stops of reverse transcriptase (lanes 1 and 2). The concentration of DMS is increased from lanes 3/4 to 5/6. Nucleotide positions are indicated on the right.*

For xY3 RNA, the MFOLD program calculated nine different structures. Only one structure, shown in Figure 3B, agreed reasonably well with the data obtained . The highly conserved sequence forming stems 1 and 2 in the human structure also form a stem structure interrupted by a small internal loop in xY3 RNA. However, the remaining part of xY3 does not seem to fold like its human counterpart. Most of the pyrimidine rich region (nts. 55-77) is predicted to be double-stranded, which is supported by the RNase V1 data. C25-C28 and A37-A40 form an internal loop structure between two small stem structures. A stem structure is formed between G43-A54 and U59-C70, containing two internal loops. The structure is supported by the probing data, but the cleavages in the same region by both RNase T1 and RNase V1 suggest that the structure is flexible.

iY3 probing – In order to obtain more information on possibly conserved secondary structures, iY3 RNA was also subjected to probing with RNase T2 (Figure 2D, lanes 4-6) and RNase V1 (Figure 2D, lanes 2 and 3). RNase V1 cleavages indicate seven distinct double-stranded regions of which three were exclusively cleaved by RNase V1 (around A16, C45, and 78A). The other four regions also show some RNase T2 cleavages. RNase T2 cleaved at two regions exclusively, around A40 and G62. The probing data for iY3 RNA are summarized in Figure 3B (right panel). As for hY3 RNA, we used MFOLD to calculate a number of secondary structures and compared them to the enzymatic probing data. For iY3 RNA, MFOLD generated eleven structures. Three computed structures were compatible with the probing results. The structure shown in Figure 3B has the lowest calculated free energy and is thus considered the most stable structure. Similar to hY3 and xY3, the structure is probably dynamic with the exception of the conserved stems 1 and 2.

In general, the most conserved secondary structure elements among the Y3 RNAs are stem 1 and stem 2. Stem 1 is identical in all three Y3 RNAs probed. There is some variation in the size of the internal loop 1 and there are minor differences in stem 2. The iguana and human structures are quite similar and show a large internal loop 2 and stem 3 structure. The Frog secondary structure, however, shows a different upper part of the molecule. Loop 2 is forming large stem structures and stem 3 shows a number of single-stranded regions. The probing data also indicate that a flexible structure of the central region is a conserved feature.

Y4 RNA

Sequences of six Y4 RNAs have been described (Figure 1C). Three of these are incomplete.[133] The first 25 nucleotides and the 3'-region from 69-94 are highly conserved. These two regions are predicted to form stems 1 and 2 (Figure 1A).[136] Furthermore, there are two central regions displaying strong sequence conservation, nts. 41 to 45 and nts. 47 to 53. The frog sequence is five nucleotides shorter than the human sequence, while the iguana sequence is two nucleotides longer (Figure 1C). Three Y4 RNAs (from human, frog and iguana) were subjected to enzymatic probing and hY4 RNA was also analysed by chemical probing using DMS.

hY4 probing – The secondary structure of hY4 RNA was first analysed with RNases A, T2, T1, and V1. Figure 4A, shows an example of enzymatic probing using RNase V1 (lanes 2-4) and RNase A (lanes 5-7). Most regions of hY4 RNA were either recognised by RNase V1, or by RNase A. The 5' region (nts. 1-20) are mainly cleaved by RNase V1, as is the 3' region between nts. 70 and 90. This supports the existence of stems 1 and 2 (Figure 1A). Nevertheless, as was the case for hY3 RNA, also in hY4 cleavages by both single-strand and double-strand specific enzymes were found in the same regions of the molecule. The hY4 RNA probing data are summarized in Figure 5A.

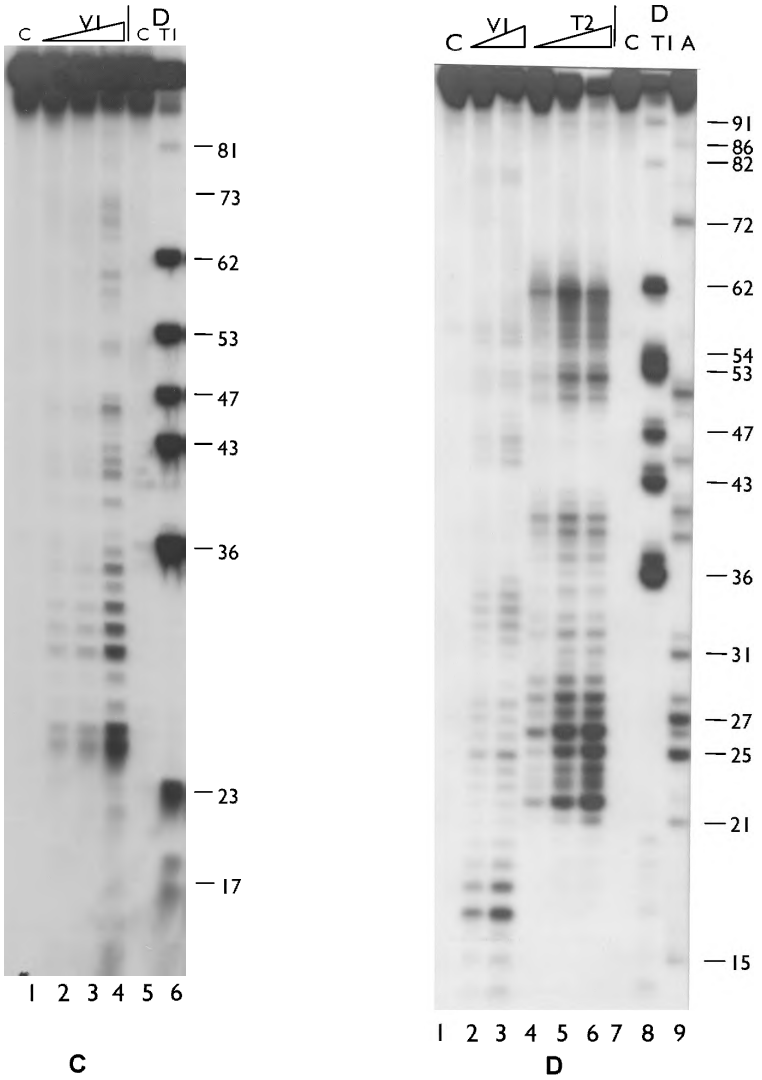


Figure 2 (Cont.). *(C) Probing experiment on xY3 RNA, using RNase V1 (lanes 2-4, increasing amounts of enzyme). Lane 1 is the control incubation in which the enzyme was omitted. Lanes 5 and 6 show a sequence ladder obtained using RNase T1 (lane 6), under denaturing conditions. Lane 5 is a control incubation. On the right the nucleotide numbering is indicated. (D) Probing experiment on iY3 RNA, using RNase V1(lanes 2 and 3, increasing amounts of enzyme) and RNase T2 (lanes 4-6, increasing amounts of enzyme). Lane 1 is the control incubation in which the enzyme was omitted. Lanes 7-9 show a sequence ladder obtained using RNase T1 (lane 8) and RNase A (lane 9), under denaturing conditions. Lane 7 is a control incubation. On the right the nucleotide numbering is indicated.*

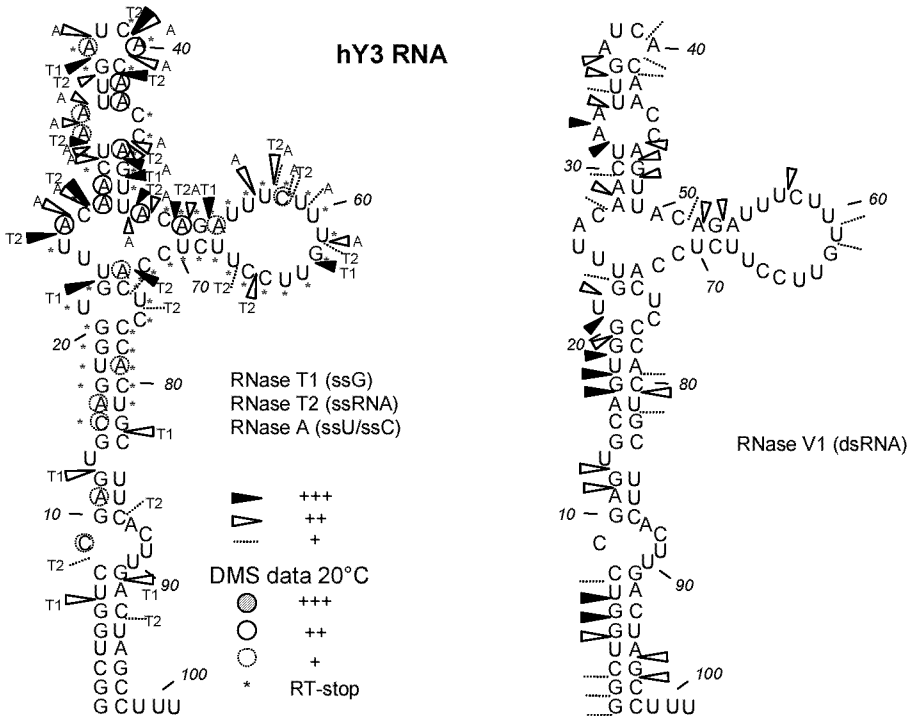


Figure 3. Summary of probing data. *(A) Summarised results of enzymatic probing on hY3 RNA, shown in the secondary structure that matches the enzymatic cleavage pattern best. In the left panel, single-stranded specific enzyme cleavages are shown (RNase T1, RNase T2, and RNase A). The right panel shows double-stranded specific cleavages by RNase V1. Dotted lines represent weak cleavages, while stronger cleavages are indicated by triangles (see legend). Also shown in the left panel are the summarised results of DMS probing on hY3 RNA at 20°C. Intensity of modification is indicated by circles (thin line: weak; bold line: average; filled circle: strong). Asterisks indicate stops of reverse transcriptase occurring in the control incubations. The oligo used for reverse transcription is complementary to the region 82-101.*

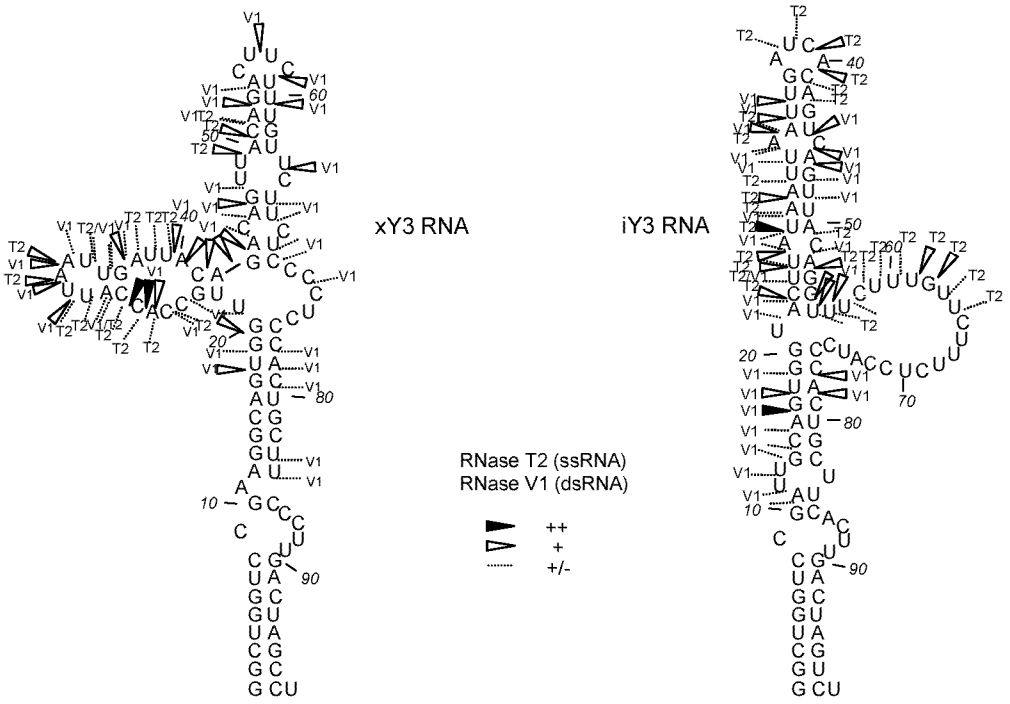


Figure 3 (Cont.). *(B) Summarised probing data for iY3 (right panel) and xY3 (left panel). RNase T2 and RNase V1 cleavages are indicated. Dotted lines represent weak cleavages, moderate cleavages are indicated by open triangles, while stronger cleavages are indicated by filled triangles (see legend).*

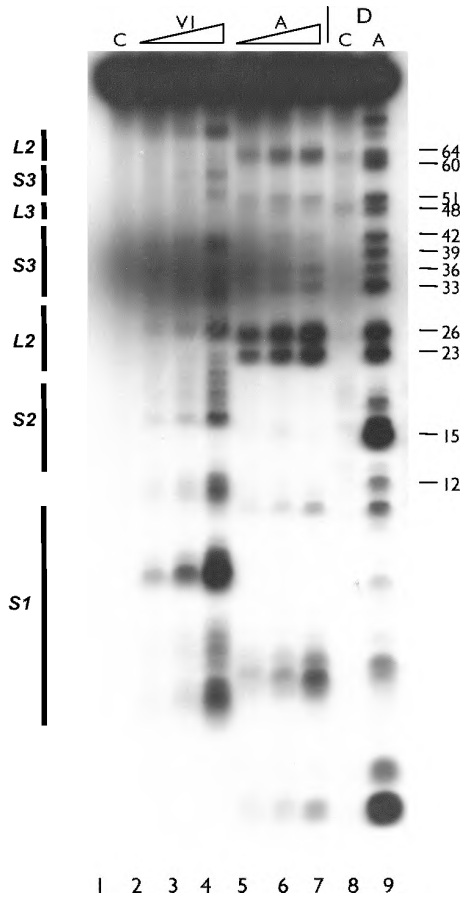


Figure 4. Secondary structure probing experiments on Y4 RNA. *(A) Probing experiment on hY4 RNA, using RNase V1 (lanes 2-4, increasing amounts) and RNase A (lanes 5-7, increasing amounts). Lane 1 is a control incubation where RNases were omitted. Lane 8 is a control incubation and lane 9 is an RNase A incubation both under denaturing conditions to generate a sequence ladder. Nucleotide positions are indicated on the right. Note that the RNase V1 cleavage products lack the 3' phosphate group and therefore migrate slightly slower though the gel.*

The MFOLD program generated nine distinct secondary structures. Four of these were compatible with the probing data. The structure, which fitted best with the experimental data also had the lowest calculated free energy and is shown in Figure 5A. There are minor changes in the structure shown in Figure 5A compared to the previously proposed structure (Figure 1A). These changes concern loop 1, which is reduced in size and changed into a bulged loop, while a small internal loop is replacing the bulged C9 residue.

Human Y4 RNA was also subjected to chemical probing using DMS, both at 20°C and at 0°C to minimise potential breathing of stem structures. Figure 4B shows the results of such an experiment. The first two lanes are control incubations to detect spontaneous stops of reverse transcriptase, which are also highly prominent for this RNA. Although these stops prevent the derivation of data for many positions, the remaining modifications support the proposed secondary structure. In stems 1 and 2 (nts. 1-21/ 70-94) a weak modification of nucleotide A73 was observed. The large

internal loop was accessible to DMS, as illustrated by modification of almost all adenosines and cytosines in the loop. Stem 3 displayed weak modifications at some A-U base pairs. The bulged A54 was accessible and seemed to open the helical structure, since A53 was also modified. The DMS data are summarised in Figure 5A. At 0°C, all nucleotides were less intensively modified compared to 20°C. A few modifications, however, were significantly diminished at the lower temperature. A24 was not modified at all, while A29-C31, A63, and A65 showed reduced modification efficiencies, suggesting that the accessibility of bases in the loop structure is reduced at the lower temperature, possibly by base-pairing.

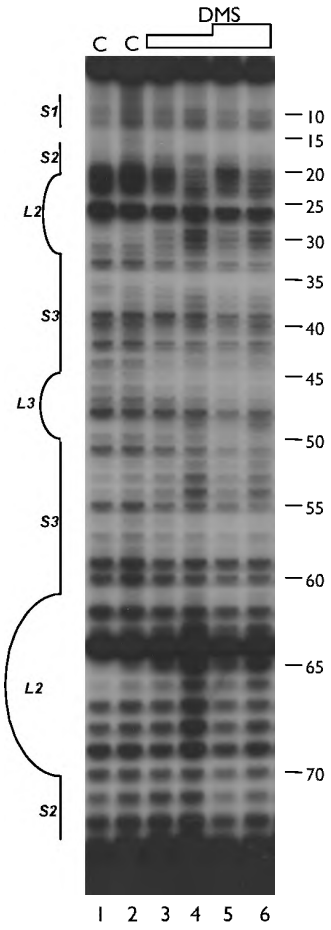


Figure 4 (Cont.). *(B) Probing hY4 RNA with DMS at 0°C (lanes 1, 3, and 5) and 20°C (lanes 2, 4, and 6). Sites of modification were detected using primer extension. Parallel control incubations without DMS were used to detect stops of reverse transcriptase (lanes 1 and 2). The concentration of DMS is increased from lanes 3/4 to 5/6. Nucleotide positions are indicated on the right.*

xY4 probing – Figure 4C shows an example of probing xY4 RNA using RNase V1 (lanes 2-4) and RNase T2 (lanes 5-7). Strong RNase V1 cleavages were found in the region up to U23, supporting the formation of the conserved stem structure. Next, a small region shows single-stranded cleavages (24-30), in part overlapping with strong cleavages by RNase V1 (27-32). In the rest of the molecule, two more single-stranded regions are found: U43-U45 and U54-U58. The data are summarised in Figure 5B (left

panel), which displays the only one of the eleven MFOLD structures that matched with the probing data.

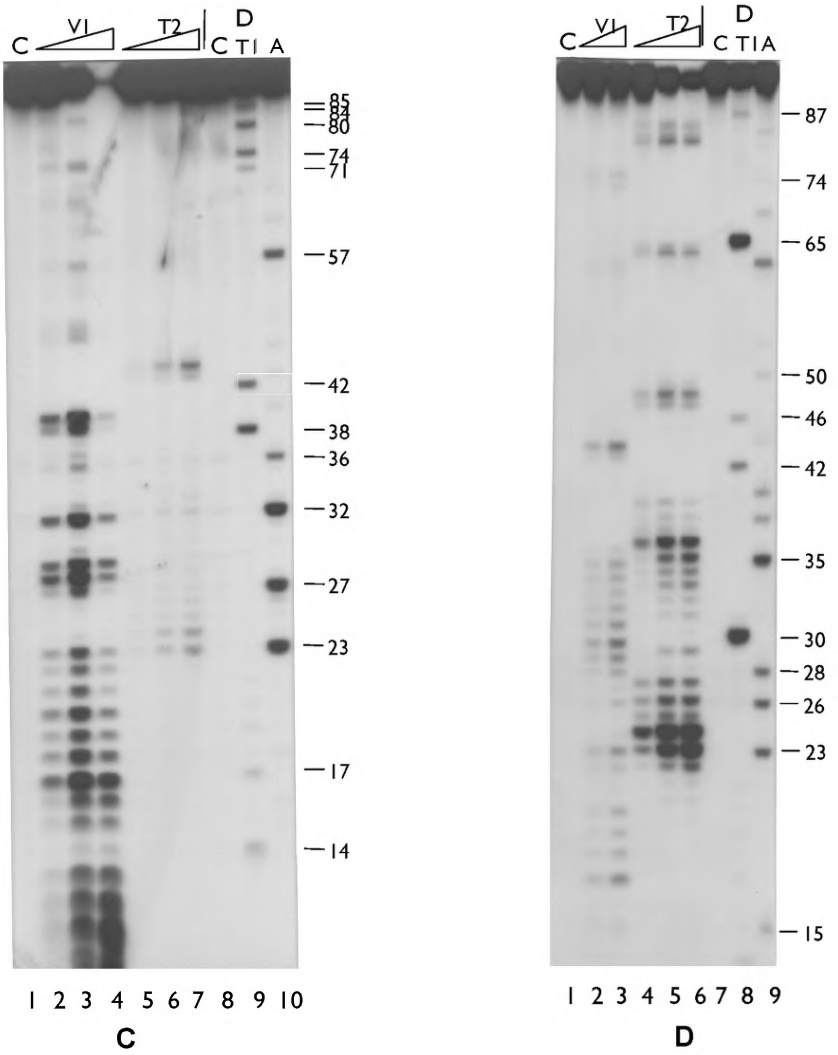


Figure 4 (Cont.). *(C) Probing experiment on xY4 RNA, using RNase V1 (lanes 2-4, increasing amounts of enzyme) and RNase T2 (lanes 5-7, increasing amounts of enzyme). Lane 1 is the control incubation in which the enzyme was omitted. Lanes 8-10 show a sequence ladder obtained using RNase T1 (lane 9) and RNase A (lane 10), under denaturing conditions. Lane 8 is a control incubation. On the right the nucleotide numbering is indicated. (D) Probing experiment on iY4 RNA, using RNase V1(lanes 2 and 3, increasing amounts of enzyme) and RNase T2 (lanes 4-6, increasing amounts of enzyme). Lane 1 is the control incubation in which the enzyme was omitted. Lanes 7-9 show a sequence ladder obtained using RNase T1 (lane 8) and RNase A (lane 9), under denaturing conditions. Lane 7 is a control incubation. On the right the nucleotide numbering is indicated.*

iY4 probing – The iguana Y4 RNA was subjected to RNase T2 and RNase V1 probing, as shown in Figure 4D. Strong RNase T2 cleavages were found between 23 and 40, while the strongest RNase V1 cleavages are found between 27 and 34. The data, summarised in Figure 5B (right panel), support three of the seven secondary structures generated by MFOLD. The structure with the lowest calculated free energy is shown in Figure 5B. In contrast to hY4 RNA, the iguana structure shows a relatively large internal loop separating stems 1 and 2 (loop 1, Figure 1A). The second internal loop is similar to, though somewhat smaller than loop 2 of the human RNA. In contrast

to the human Y4 RNA, stem 3 of iY4 RNA contains a large internal loop. Interestingly, this secondary structure resembles that of xY4 RNA.

In spite of the differences the three Y4 RNA molecules analysed showed conserved structural features. Stems 1, 2, and 3 are formed in all three RNAs. The internal loop between stems 1 and 2 is reduced in size in hY4 compared to iY4 and xY4 RNA. The reverse is true for internal loop 2, which is smaller in iY4 and xY4 RNA than in hY4 RNA. Finally, stem 3 is interrupted by a relatively large internal loop in xY4 and iY4 RNA.

Discussion

In this study the secondary structures of three Y3 and three Y4 RNAs from different organisms were studied in order to gain insight in the conservation of structural elements. The Y3 and Y4 RNAs from human, iguana, and frog were analysed by enzymatic and chemical probing. As expected the data fully support the formation of the long stem (stem 1 + stem 2) by base pairing of the 5'- and 3'-ends of the RNA in all of these molecules. This region, which contains the most highly conserved nucleotides, not only among homologous Y RNAs from different species but also among different Y RNAs, contains the binding site for the Ro60 protein [127,130,131] as well as an element important for nuclear export of the Y RNAs.[175] More importantly, the present data demonstrate that the less well conserved central parts of the Y3 and Y4 RNAs do not fold into very stable structural elements under physiological conditions, suggesting that these parts of the molecules are rather dynamic and may transiently adopt various alternative secondary structures. Interestingly, this phenomenon does not seem to be restricted to the Y3 and Y4 RNAs, but is probably also a characteristic feature of the Y1 and Y5 RNAs, since 'breathing' of the central regions of the latter RNAs has been proposed previously based upon the results from similar probing experiments with the human Y1 and Y5 RNA.[136] Therefore it is highly likely that the central region of all Y RNAs is dynamic in nature and thus is readily available for base pairing interactions with another nucleic acid, which might be important for its function. The stable base paired stem of hY3 RNA (stem 1 and stem 2) might serve as a 'handle' for proteins (e.g. Ro60), while the central part might function as a 'fishing net' to catch other molecules based upon its capacity to form (transient) intermolecular base pairs. An attractive, though still rather speculative, target for such an interaction is the family of 5'TOP mRNAs, a family of mRNAs encoding proteins, most, if not all of which play some role in the protein synthesis machinery, as for example ribosomal proteins. It has been established that these proteins are co-ordinately expressed and that the 5'UTRs of the respective mRNAs play an important role in the regulation of this process. Recently, two factors interacting with such 5'UTRs have been identified, CNBP and the La protein.[176,177] In addition the Ro60 as well as a yet unidentified RNA have been demonstrated to be required for these interactions.[126] Therefore it is tempting to speculate that a Ro60- (and La-) associated Y RNA is involved in the regulation of translation of 5'TOP mRNAs and that selectivity is in part introduced by a

transient base pairing interaction between the central region of a Y RNA and the 5'UTR of the mRNA.

In spite of the apparent structural heterogeneity in the central regions of the Y RNAs, the experimental structure probing data for each of the Y RNAs analysed were compared with the most stable secondary structures predicted by computer algorithms to establish which of the predicted structures most well corresponded to the biochemical results. In comparison with the previously proposed secondary structure for hY3 RNA (Figure 1A) the probing experiments of hY3 RNA support an alternative structure in which the large internal loop 2 (Figure 1A) is in part replaced by a small stem structure (Figure 3A) and in which loop 1 is somewhat shifted. For the 'upper' part of the molecule, including loop 2, ambiguous enzymatic probing data were obtained, indeed indicating that the structural elements that can be formed in the central part of hY3 RNA are relatively unstable and that this part of the molecule may adopt several alternative structures. Probing experiments performed at 0°C, a temperature at which base pairing interactions are stabilised, showed the expected general stabilisation of the RNA structure and provided additional support for the structure shown in Figure 3A. Interestingly, in at least one of the hY3 RNA structures formed in solution a pseudoknot may be formed in the central part of the molecule. A pseudoknot in hY3 is not only supported by enzymatic probing data, but also by the conservation of this part of the sequence between the human and mouse RNAs. Furthermore, all nucleotide substitutions in the corresponding sequence elements of iguana, pig, dog, and cow Y3 RNA allow pseudoknot formation (see Figure 1B). *Xenopus* Y3 is the only RNA that does not seem to be able to form a corresponding pseudoknot.

The probing results for xY3 RNA are in agreement with previously published data.[130] While, like in the other Y RNAs, the formation of stems 1 and 2 and internal loop 1 are evident, the structure of the central region of the most likely xY3 RNA is quite different from that of the human and iguana structures. Stem 3 seems to be formed by nucleotides 43-54 and 59-70. This implies that the pyrimidine rich region around position 60 is part of a double-stranded region, as opposed to being single-stranded in hY3 and iY3. A smaller pyrimidine-rich region comprising C71-C75 is, however, single-stranded. Compared to the human structure, the iY3 probing data suggest that the basic structure is identical. The first two stem structures are formed, placing C9 in an internal loop, similar to the situation in hY3 RNA. In iY3 RNA there is an additional internal loop at a position where the human structure shows a bulged residue (U13). Stem 3, formed by A22-G36 and C41-U55, is longer in iY3 than in hY3 RNA. This stem is interrupted by two mismatches and contains a large number of low stability base pairs, A-U and G-U. In contrast to the hY3 RNA structure, the loop 2 region is completely single-stranded, in agreement with the RNase T2 data (Figure 3B). As observed for hY3, the pyrimidine-rich region in iY3 (U63-C75) was hardly accessible for ribonucleases, although it is predicted to be single-stranded.

Also for hY4 RNA the most likely secondary structure was derived from a combination of enzymatic and chemical probing data and computer generated

secondary structure variants. The resulting structure (Figure 5A) has only minor changes compared to the previously proposed structure (Figure 1A). Loop 1 was changed into a bulged loop and C9 is now part of a mismatch. This change is supported by RNase V1 data for the region around G10, suggesting that these nucleotides adopt a helical conformation. On the opposite side, around A80, the non-base paired residues appeared to be accessible for single-strand specific enzymes. RNase V1 only cleaved inefficiently around U84, which indicates that either this mismatched residue stacks into the helix or that a minority of molecules adopts a different conformation, like for instance the structure depicted in Figure 1A. The remainder of hY4 RNA is identical to the previously proposed structure. hY4 probing experiments using DMS at 0°C indicated a general stabilisation of the RNA structure. The accessibility of the internal loop 2, however, seemed to be more reduced than the rest of the RNA, suggesting that the loop may fold into a more stable structure.

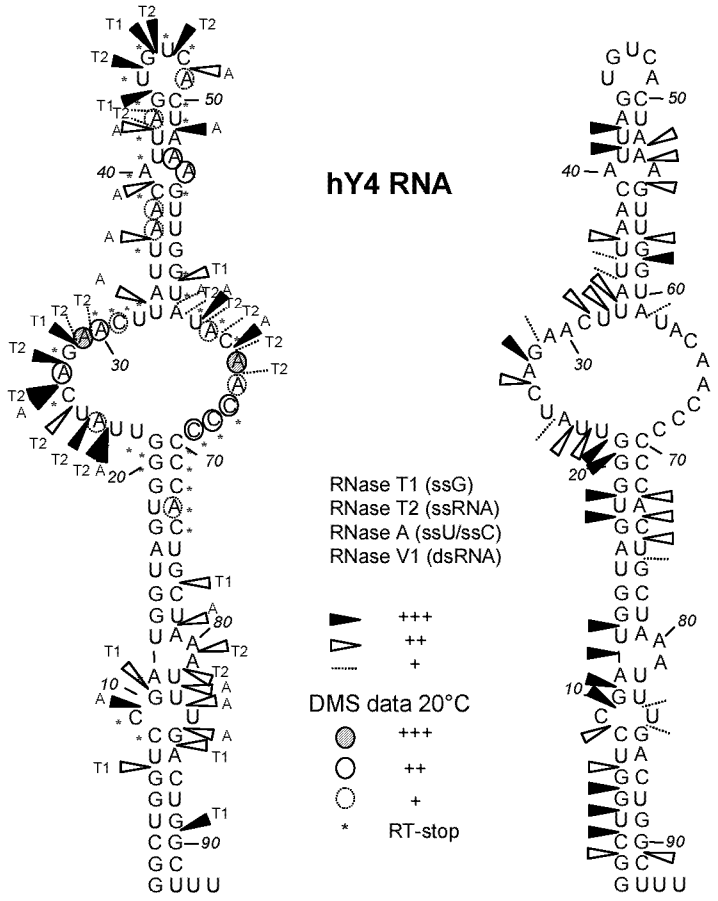


Figure 5. Summary of probing data. *(A) Summarised results of enzymatic probing on hY4 RNA, shown in the secondary structure that matches the enzymatic cleavage pattern best. In the left panel, single-stranded specific enzyme cleavages are shown (RNase T1, RNase T2, and RNase A). The right panel shows double-stranded specific cleavages by RNase V1 (in black). Dotted lines represent weak cleavages, while stronger cleavages are indicated by triangles (see legend). Also shown in the left panel are the summarised results of DMS probing on hY4 RNA at 20°C. Intensity of modification is indicated by circles (thin line: weak; bold line: average; filled circle: strong). Asterisks indicate stops of reverse transcriptase occurring in the control incubations. The oligo used for reverse transcription is complementary to the region 82-101.*

The most likely secondary structures derived for iY4 and xY4 are highly similar, but differ somewhat from the hY4 RNA structure. While internal loop 1 is somewhat larger, loop 2 is smaller in both iY4 and xY4. In addition, stem 3 contains an internal loop in the iguana and frog structure rather than a mismatch as in hY4 RNA.

Taken together the major differences between the most likely secondary structures of Y3 and Y4 RNAs from different species were found in the central parts of the molecules. This substantiates the idea that the flexibility rather than the structure of this region of Y RNAs is of functional importance and that there is no or only a very limited evolutionary selection pressure on the structural elements that are present in the central parts of the most likely secondary structures.

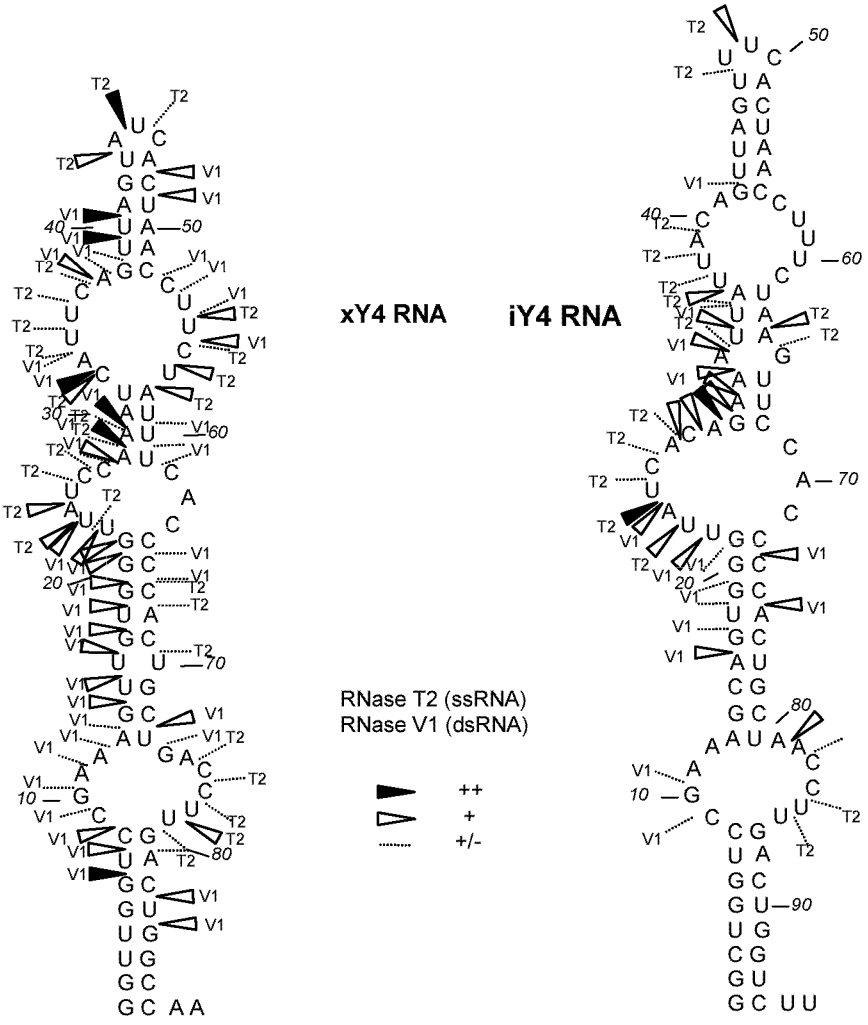


Figure 5 (Cont.). *(B) Summarised probing data for iY4 (right panel) and xY4 (left panel). RNase T2 and RNase V1 cleavages are indicated. Dotted lines represent weak cleavages, moderate cleavages are indicated by open triangles, while stronger cleavages are indicated by filled triangles (see legend).*

Recently, a novel RNA secondary structure comparison algorithm, the suboptimal RNA analysis program (SORA), has been developed and has been used to find phylogenetically conserved secondary structure models for the Y1, Y3 and Y4 RNAs.[173] The biochemical data obtained in the present study allow a validation of the Y3 and Y4 SORA solution structures. Most interestingly, the SORA solution structures for the stem 1 – stem 2 region of hY3 and hY4 RNA are fully supported by the results of the probing experiments. In addition, SORA predicted the presence of a relatively large internal loop in both Y3 and Y4 RNA in the loop 2 region, which is in agreement with the absence of a very stable secondary structure element in this region of these molecules. Finally, the structure of the stem 3 region predicted by SORA does not or only partially correspond to that derived in the present study. However, we should stress that the experimental probing data for this region are rather ambiguous and, as noted above, might be the cumulative results for a number of several alternative structures that exist in solution under the conditions of the probing experiments. Taken together, these results indicate that the SORA solution structures for the Y RNAs correlate well with the experimental probing data, which suggests that the SORA method provides a reliable tool for RNA secondary structure prediction, in particular for regions in RNA molecules that do not fold into relatively stable secondary structural elements.

Since all structure probing experiments were performed *in vitro* in the absence of the proteins that associate with the Y RNAs in cells, we can not exclude that protein binding may have some effect on the structure of the Y RNAs. In fact, it has recently been shown that subtle structural changes may occur in stem 1 – loop1 – stem 2 region of xY3 RNA upon binding of the Ro60 protein.[130] In addition to structural changes caused by protein binding, a bound protein may also stabilise a particular structural element that, in the absence of protein, exists in equilibrium with an alternative structure. If such phenomena occur in the dynamic central region of Y RNAs awaits the identification and characterisation of proteins binding to this region, which may not exist at all.

We have obtained experimental evidence for conserved structural features between Y RNAs of different organisms. It was shown that the most highly conserved sequences also form conserved structural elements. Furthermore, we found that a flexible part of the RNA is also a conserved feature (rather than a conserved structure). This might indicate which regions are important for the function of the molecule. Future studies at the three dimensional level will ultimately give evidence for structurally conserved elements and show which nucleotides are important for structure formation and hence, for function.

Acknowledgements

We thank Dr. S. Wolin (Yale University School of Medicine, New Haven, CT, USA) for kindly providing *Xenopus* Y3 and Y4 RNA constructs. This work was supported

in part by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO).

Materials and Methods

Genbank accession numbers. The Y RNA sequences can be accessed through the following codes: human Y3, K01563, frog Y3 L15431, mouse Y3 U34827, iguana Y3 L27532, cow Y3 U84671, duck Y3 U82125, pig Y3 U84674, dog Y3 U84672, human Y4 L32608, frog Y4 L15432, iguana Y4 L27537, dog Y4 U84669, cow Y4 U84668, monkey Y4 U84670.

In vitro transcription of RNA. The hY RNAs are cloned in the pUC19 vector, containing a *Dra*I site. Linearisation with *Dra*I, followed by T7 transcription,[168] results in RNA molecules without any additional vector derived nucleotides (as shown in Figure 1). The frog Y3 and Y4 constructs were a kind gift from dr. S. Wolin. The frog and iguana Y RNA encoding sequences were isolated in a PCR reaction using oligonucleotides containing a T7 promoter site and a *Dra*I site. The constructs were subcloned in pUC19, sequenced and, after *Dra*I digestion, used for transcription.

5'- and 3'-endlabelling. Three different methods of endlabelling were used in the experiments. 3'-Endlabelling was achieved via ligating [^{32}P]-pCp using T4 RNA ligase (Amersham Pharmacia Biotech) for one hour at 37 °C. 5'-Endlabelling was either performed as described,[168] or during the *in vitro* transcription using 50 μCi [γ-^{32}P]-GTP. The reaction mixture contained 1 μg DNA template, 0.01 μg BSA, 0.01 mM DTE, [γ-^{32}P]-GTP and T7 RNA polymerase in a total volume of 50 μl. After a five minutes incubation at 37 °C, 0.1 mM of each (unlabelled) nucleotide was added and incubated for one hour. Each of the labelled products was isolated from a 10% PAA/8M urea gel.

RNA structure probing. The RNA molecules were subjected to enzymatic cleavage by RNase V1 (dsN, $7x10^{-4}$ U/μl, Amersham Pharmacia Biotech), A (ssC & ssU, $1x10^{-6}$ U/μl, Amersham Pharmacia Biotech), T1 (ssG, 2 U/μl, Amersham Pharmacia Biotech), and T2 (ssN, $4x10^{-3}$ U/μl, Gibco BRL), under single hit conditions and at 20°C. In all experiments, parallel incubations in which the enzyme was omitted, served as a control for spontaneous RNA breaks. Reaction conditions are essentially as described by.[168] Probing under denaturing conditions (8 M urea, 1 mM EDTA, 50°C.) were performed in order to obtain a sequence ladder.

Chemical probing using DMS, followed by primer extension to determine sites of modification, was performed as described previously.[168] The oligos used in the primer extension reactions were complementary to the following regions: 82-101(hY3) and 76-94 (hY4).

Secondary structure predictions and sequence alignments. The secondary structure predictions were done with MFOLD version 3.0. The temperature for the predictions was set to 20°C, the percent suboptimality to 50. The sequence alignments were made using the CLUSTALW V1.7 program [178] and adjusted manually.

5

General Discussion

The last decades have shown an increased understanding of RNA structure and function. Numerous structural building blocks have been discovered and, in some cases, characterised in atomic detail. These RNA structural elements show a diversity and complexity comparable to that of protein structures. In fact, RNA structure can, although limited to only four basic elements, be just as versatile as protein structure. RNA can accomplish this through unusual base pairing, by base modifications, and using secondary structure building blocks that can be combined to a complex tertiary folding. Several studies have unequivocally demonstrated the importance of tertiary folding for the function of RNAs, as for example in the case of catalytic RNAs, which substantiates the need for detailed structural analyses. Unusual and/or interrupted base pairing causes distortion of regular helical structures leading to the exposure of different chemical groups to the surroundings. In regular structures, e.g. the double helix, sequence specific recognition is difficult because the characteristic chemical groups are involved in base pairs formation and are therefore not readily accessible. Thus, unusual base pairing enhances specific recognition of RNA sequence by either proteins or nucleic acids. Base modifications that take part in tertiary structure formation have been known since the first determination of the X-ray structure of $tRNA^{Phe}$.[4] The importance of base modifications is similar to that of unusual base pairing. Different chemical groups facilitate specific interactions with either other parts of the RNA itself or with other molecules. The secondary structure elements (double helix, single-stranded, hairpin loops, tetraloop receptor, etc) can be combined with the modifications and unusual base pairings to form a unique tertiary conformation. This unique three-dimensional folding determines whether and how efficient the RNA will be able to perform its function.

Structural flexibility of RNA vs. structure probing

One of the most important characteristics of RNA, which distinguishes RNA from DNA, is the flexibility of the tertiary structures. As discussed in Chapter 1, for example, large structural rearrangements take place when the tetraloop binds to its receptor. In biological systems, multiple conformations can be of functional relevance for an RNA molecule. Different conformations can be accomplished within a single RNA molecule (RNA dynamics): 1) rearrangements of entire secondary structure elements can create an entirely different folding; 2) changes within secondary structure elements (e.g. single-stranded regions combine into a double stranded region); 3) alterations of tertiary interactions. It is also possible that several populations of different stable conformations coexist: folding redundancy.

Probing of RNA structure, using both nucleolytic enzymes and chemical probes, can increase the knowledge on the dynamics and/or folding redundancy of RNA. The bulky size of enzymes has the disadvantage that steric hindrance is a major factor to be taken into consideration. This means that certain cleavages will not occur because

the active site of the enzyme can not physically interact with the cleavage site. For example, RNase V1 cleaves only double-stranded regions if it can interact with at least five nucleotides which have a double-stranded conformation. On the other hand, enzymes are more sensitive in detecting changes in tertiary folds than most of the chemical reagents used in such studies. If such changes are for instance caused by protein binding, while the secondary structure remains unchanged, chemical reagents specific for secondary structure determination will show similar results, because they are small and are able to reach bases 'buried' within the molecule.

If a difference is found in an enzymatic cleavage pattern obtained under different experimental conditions (e.g. different Mg2+ concentrations, different incubation temperatures, or probing in the presence and absence of a binding protein), this difference can originate from several factors as discussed above. The information obtained from these experiments is just the presence or absence of enzyme cleavage at certain positions in the RNA and the additional appearance or disappearance of those cleavages that occur at only one of the experimental conditions. As a consequence, it will be difficult to discriminate between, for example, rearrangements of tertiary folding with the same secondary structure and rearrangements of the secondary structure itself. A combination of the two effects, which is more likely, will result in a cleavage pattern that mostly cannot be interpreted unambiguously. Bands that appear when a single-strand specific enzyme is used may indicate that the region becomes single-stranded. On the other hand, it could be that the same region already was single-stranded but not accessible to the enzyme. It is therefore important to perform such experiments under a number of different conditions (different buffers, temperatures, RNA concentrations, ionic conditions, etc.) and with a number of different reagents.

Additionally, the use of chemical reagents, of which some provide secondary structure information (like DMS and CMCT), aids in distinguishing whether changes in enzymatic cleavage patterns are the result of secondary structure alterations or of changes in tertiary structure. Reagents which are indicative for tertiary interactions (like DEPC, Fe(II)EDTA, DMS or ENU) help in detecting tertiary rearrangements of the RNA structure.

Probing experiments can yield a wealth of structural information on naked RNA and on RNA interacting with other molecules. When used and interpreted correctly, these relatively simple experiments can provide important information on the relationschip between the structure and the function of the RNA.

Preferably, structure probing data should be supported by experimental results obtained with a different approach, either biochemical, such as crosslinking or mutational studies, or biophysical (NMR, X-ray diffraction). The aim of probing experiments should not *per se* be to obtain a single secondary structure, but to examine and determine the whole range of structures formed by the RNA, including regions that display some flexibility and/or structural redundancy and regions that remain stable under different conditions. In this way a general impression of the structure of the RNA of interest might provide clues for its function in vivo.

The U1A-3'UTR complex

The U1A protein is one of the best characterised RNA binding proteins. One of the two RNP motifs, binds to two different RNA targets. Its primary binding sequence is found in U1 snRNA which is part of U1 snRNP and plays an important role in the splicing of pre-mRNA. In addition, it binds to the 3'utr of its own pre-mRNA, is bound only when the expression level of U1A protein is sufficiently high to bind all U1 snRNA molecules. The remaining U1A protein then binds to the two recognition sequences found in the 3'utr (one identical to the U1 snRNA binding sequence and one containing a single nucleotide substitution) which inhibits further processing of the pre-mRNA and causes subsequent down regulation of the protein expression level. Several studies, including X-ray diffraction and NMR, on the U1A – U1 snRNA complex and on the U1 snRNA have shown that the binding is an intricate and complex mechanism. When the U1A protein is bound to its target, not only the protein shows large structural rearrangements,[117] but also the RNA changes its three-dimensional structure in order to accommodate specific amino acid - base interactions. Less structural information is available on the complex of U1A bound to its pre-mRNA. The experimental data presented in chapter 2 indicated two proteins bound on the same side of the RNA. Enzymatic probing indicated a large region being protected by the proteins. The results also indicated a structural change in the RNA upon protein binding demonstrating flexibility in RNA molecules which is necessary for binding and hence for function. Another elegant structure probing study provided additional information on the RNA-protein interaction, by introducing a tethered FE(II)EDTA group in the U1A protein.[83] By this approach the orientation of the protein on the RNA could be determined. The combined studies provide a global view on the RNA-protein complex. These results were later confirmed by NMR data and other studies.[118,124,125] The complementarity of the data obtained by the different techniques demonstrates the power of biochemical RNA structure probing techniques in the absence of crystallographic and/or NMR data.

The U1 snRNA - scFv complex

Patients suffering from systemic lupus erythematosus can produce autoantibodies against U1 snRNP. The antigenic regions in the RNA were previously mapped to areas where no protein is bound: Stem II and stem-loop IV.[153-155] These studies were performed using sera of patients that contain multiple RNA binding activities. The isolation of a recombinant monoclonal autoantibody fragment (presented in Chapter 3), allowed the study a single RNA binding site or epitope on the U1 snRNA. Two important binding characteristics were found. The first one was that there is a conformational change in the RNA to accommodate binding. Upon binding of the scFv, the RNA double-stranded region was stabilised, as were the most 3' three nucleotides of the loop, which are probably stacked on each other in the protein-RNA complex.

The detailed chemical probing of the scFv – U1 snRNA complex, using Fe(II)EDTA as probe, demonstrated the second important binding characteristic. The experiments showed that a distorted part of the helical structure forms the heart of the binding site of the monoclonal antibody fragment.[179] The unusual, non-canonical base pairing in this region, which is stabilised by surrounding regular helices, exposes the bases and represents most likely the most important determinant for the sequence specific recognition by the scFv.

Y RNAs

The examples above show induced fit (or accommodated binding) between RNA and protein. The RNA itself, without bound protein, can also be very flexible. As can be seen for the Y RNAs, multiple conformations exist in solution. The Y RNAs, especially hY1 and hY3, contain a large internal loop predicted to be single stranded. The conformation of this element is highly dependent on the magnesium ion concentration. In the absence of magnesium, loop 2B (see Figure 8, Chapter 1) of hY1 is readily accessible to probing reagents. If, however, the concentration of the magnesium ions is increased, the loop structure becomes gradually less accessible to enzymatic and chemical probes.[136] This indicates a large structural change within the RNA molecule. It is tempting to speculate that this loop, dependent on the ionic conditions, is interacting with different regions of the RNA. Furthermore, the probing experiments presented in Chapter 4 indicate that these regions in Y RNAs exist as several alternative three-dimensional folds. In other words, stretches of nucleotides are readily interacting with different parts of the RNA, reflecting its flexible, dynamic nature. It is likely that this flexibility is the key property for the RNA to perform its (possible) function in binding to other nucleic acids such as the 5'TOP mRNAs.[126] From this perspective, the probing experiments also may provide clues to which regions of the RNA are important to perform its function, which may be a starting point for new studies.

In general, RNA structure probing can be a powerful technique to obtain detailed information on RNA structure, dynamics, and different conformations, both for the naked RNA, as well as for RNA in RNA-protein complexes. RNA structure probing is an excellent alternative when NMR or crystallographic data are not available and provides useful complementary information when such data are available.

References

1. Chastain, M.; Tinoco, I. ,Jr. *Prog. Nucleic Acid Res. Mol. Biol.* 1991, *41*:131-177
2. Strobel, S. A.; Doudna, J. A. *Trends Biochem. Sci.* 1997, *22*, 262-266
3. Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* 1996, *273*, 1678-1685
4. Robertus, J. D.; Ladner, J. E.; Finch, J. T.; Rhodes, D.; Brown, R. S.; Clark, B. F.; Klug, A. *Nature* 1974, *250*, 546-551
5. Ladner, J. E.; Jack, A.; Robertus, J. D.; Brown, R. S.; Rhodes, D.; Clark, B. F.; Klug, A. *Proc. Natl. Acad. Sci. U. S. A.* 1975, *72*, 4414-4418
6. Holbrook, S. R.; Cheong, C.; Tinoco, I. ,Jr.; Kim, S. H. *Nature* 1991, *353*:579-581
7. Leonard, G. A.; McAuley-Hecht, K. E.; Ebel, S.; Lough, D. M.; Brown, T.; Hunter, W. N. *Structure* 1994, *2*:483-494
8. Cruse, W.; Saludjian, P.; Biala, E.; Strazewski, P.; Prange, T.; Kennard, O. *Proc. Natl. Acad. Sci. U. S. A.* 1994, *91*:4160-4164
9. Betzel, D.; Lorenz, S.; Furste, J. P.; Bald, R.; Zhang, M.; Schneider, T. R.; Wilson, K. S.; Erdmann, V. A. *FEBS letters* 1994, *351*:159-164
10. Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Moras, D.; DeBear, J. S.; Gough, G. R.; Gilham, P. T.; Johnson, J. *Nature* 1988, *335*:375-378
11. Lietzke, S. E.; Barnes, C. L.; Kundrot, C. E. *Curr. Opin. Struct. Biol.* 1995, *5*, 645-649
12. Pley, H. W.; Flaherty, K. M.; McKay, D. B. *Nature* 1994, *372*:68-74
13. Scott, W. G.; Finch, J. T.; Klug, A. *Cell* 1995, *81*:991-1002
14. Doudna, J. A.; Cate, J. H. *Curr. Opin. Struct. Biol.* 1997, *7*, 310-316
15. Conn, G. L.; Draper, D. E. *Curr. Opin. Struct. Biol.* 1998, *8*, 278-285
16. Gautheret, D.; Konings, D.; Gutell, R. R. *RNA* 1995, *1*, 807-814
17. Gautheret, D.; Konings, D.; Gutell, R. R. *J Mol. Biol.* 1994, *242*, 1-8
18. Walczak, R.; Westhof, E.; Carbon, P.; Krol, A. *RNA* 1996, *2*, 367-379
19. Walczak, R.; Carbon, P.; Krol, A. *RNA* 1998, *4*, 74-84
20. Auffinger, P.; Westhof, E. *J.Mol. Biol.* 1997, *274*, 54-63
21. Wu, M.; Tinoco, I. *Proc. Natl. Acad. Sci. U. S. A.* 1998, *95*, 11555-11560
22. Price, S.; Nagai, K. *Structure* 1996, *4*, 1129-1132
23. Van de Ven, F. J.; Hilbers, C. W. *Eur. J Biochem.* 1988, *178*, 1-38
24. Millar, D. P. *Curr. Opin. Struct. Biol.* 1996, *6*, 322-326
25. Walter, F.; Murchie, A. I. H.; Duckett, D. R.; Lilley, D. M. J. *RNA .* 1998, *4*, 719-728
26. Ciesiolka, J.; Yarus, M. *RNA* 1996, *2*, 785-793
27. Hagerman, P. J.; Amiri, K. M. A. *Curr. Opin. Struct. Biol.* 1996, *6*, 317-321
28. Jaeger, L. *Curr. Opin. Struct. Biol.* 1997, *7*, 324-335
29. Draper, D. E.; Misra, V. K. *Nature Struct. Biol.* 1998, *5*, 927-930
30. Cate, J. H.; Doudna, J. A. *Structure* 1996, *4*, 1221-1229
31. Cate, J. H.; Hanna, R. L.; Doudna, J. A. *Nature Struct. Biol.* 1997, *4*, 553-558
32. Pyle, A. M. *Science* 1993, *261*, 709-714
33. Ippolito, J. A.; Steitz, T. A. *Proc. Natl. Acad. Sci. U. S. A.* 1998, *95*:9819-9824
34. Lehnert, V.; Jaeger, L.; Michel, F.; Westhof, E. *Chem. Biol.* 1996, *3*:993-1009
35. Jestin, J. L.; Deme, E.; Jacquier, A. *EMBO J* 1997, *16*:2945-2954
36. Costa, M.; Deme, E.; Jacquier, A.; Michel, F. *J Mol. Biol.* 1997, *267*:520-536
37. Massire, C.; Jaeger, L.; Westhof, E. *RNA* 1997, *3*:553-556
38. Butcher, S. E.; Dieckmann, T.; Feigon, J. *EMBO J* 1997, *16*, 7490-7499
39. Basu, S.; Rambo, R. P.; StraussSoukup, J.; Cate, J. H.; FerreDAmare, A. R.; Strobel, S. A.; Doudna, J. A. *Nature Struct. Biol.* 1998, *5*, 986-992
40. Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Szewczak, A. A.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* 1996, *273*, 1696-1699
41. Basu, S.; Rambo, R. P.; StraussSoukup, J.; Cate, J. H.; FerreDAmare, A. R.; Strobel, S. A.; Doudna, J. A. *Nature Struct. Biol.* 1998, *5*:986-992
42. Heus, H. A.; Wijmenga, S. S.; Hoppe, H.; Hilbers, C. W. *J. Mol. Biol.* 1997, *271*, 147-158
43. Biswas, R.; Wahl, M. C.; Ban, C.; Sundaralingam, M. *J Mol. Biol.* 1997, *267*:1149-1156
44. Biswas, R.; Sundaralingam, M. *J Mol. Biol.* 1997, *270*:511-519
45. Brunel, C.; Romby, P.; Westhof, E.; Ehresmann, C.; Ehresmann, B. *J Mol. Biol.* 1991, *221*, 293-308
46. Wimberly, B. *Nat. Struct. Biol.* 1994, *1*, 820-827
47. Katahira, M.; Kanagawa, M.; Sato, H.; Uesugi, S.; Fujii, S.; Kohno, T.; Maeda, T. *Nucleic Acids Res.* 1994, *22*, 2752-2759
48. Katahira, M.; Sato, H.; Mishima, K.; Uesugi, S.; Fujii, S. *Nucleic Acids Res.* 1993, *21*, 5418-5424

49. Walter, A. E.; Wu, M.; Turner, D. H. *Biochemistry* **1994**, *33*, 11349-11354
50. Lane, A.; Ebel Sy; Brown, T. *Eur. J Biochem.* 1994, *220*, 717-727
51. Knapp, G. *Methods Enzymol.* **1989**, *180*,
52. Krol, A.; Carbon, P. *Methods Enzymol.* **1989**, *180*,
53. Ehresmann, C.; Baudin, F.; Mougel, M.; Romby, P.; Ebel, J. P.; Ehresmann, B. *Nucleic Acids Res.* 1987, *15*, 9109-9128
54. Lowman, H. B.; Draper, D. E. *J Biol. Chem.* 1986, *261*, 5396-5403
55. Huber, P. W. *FASEB J* **1993**, *7*, 1367-1375
56. Pogozelski, W. K.; McNeese, T. J.; Tullius, T. D. *J. Am. Chem. Soc* **1995**, *117*: 24.6428-6433
57. Berkhout, B.; Klaver, B.; Das, A. T. *Nucleic Acids Res.* **1997**, *22*:940-947
58. Porse, B. T.; Garrett, R. A. *J Mol. Biol.* 1995, *249*:1-10
59. Siebenlist, U.; Simpson, R. B.; Gilbert, W. *Cell* 1980, *20*:269-281
60. Karaoglu, D.; Thurlow, D. L. *Nucleic Acids Res.* **1991**, *19*, 5293-5300
61. Scripkin, E.; Paillart, J. C.; Marquet, R.; Ehresmann, B.; Ehresmann, C. *Proc. Natl. Acad. Sci. U. S. A.* 1994, *91*:4945-4949
62. Bassi, G. S.; Mollegaard, N. E.; Murchie, A. I.; von Kitzing, E.; Lilley, D. M. *Nat. Struct. Biol.* 1995, *2*:45-55
63. Bassi, G. S.; Murchie, A. I.; Lilley, D. M. *RNA* 1996, *2*:756-768
64. Frank, J. *Curr. Opin. Struct. Biol.* 1997, *7*:266-272
65. Frank, J.; Zhu, J.; Penczek, P.; Li, Y.; Srivastava, S.; Verschoor, A.; Radermacher, M.; Grassucci, R.; Lata, R. K.; Agrawal, R. K. *Nature* 1995, *376*:441-444
66. Stark, H.; Mueller, F.; Orlova, E. V.; Schatz, M.; Dube, P.; Erdemir, T.; Zemlin, F.; Brimacombe, R.; van Heel, M. *Structure* 1995, *3*:815-821
67. Agrawal, R. K.; Penczek, P.; Grassucci, R. A.; Li, Y.; Leith, A.; Nierhaus, K. H.; Frank, J. *Science* 1996, *271*:1000-1002
68. Stark, H.; Orlova, E. V.; Rinke Appel, J.; Junke, N.; Mueller, F.; Rodnina, M.; Wintermeyer, W.; Brimacombe, R.; van Heel, M. *Cell* 1997, *88*:19-28
69. Chen, J. L.; Nolan, J. M.; Harris, M. E.; Pace, N. R. *EMBO J* 1998, *17*, 1515-1525
70. Sontheimer, E. J. *Mol. Biol. Rep.* 1994, *20*, 35-44
71. McGregor, A.; Rao, M. V.; Duckworth, G.; Stockley, P. G.; Connolly, B. A. *Nucleic Acids Res.* 1996, *24*, 3173-3180
72. Earnshaw, D. J.; Masquida, B.; Muller, S.; Sigurdsson, S. T.; Eckstein, F.; Westhof, E.; Gait, M. J. *J. Mol. Biol.* 1997, *274*, 197-212
73. Baranov, P. V.; Dokudovskaya, S. S.; Oretskaya, T. S.; Dontsova, O. A.; Bogdanov, A. A.; Brimacombe, R. *Nucleic Acids Res.* 1997, *25*, 2266-2273
74. Ping, Y. H.; Liu, Y.; Wang, X.; Neenhold, H. R.; Rana, T. M. *RNA* 1997, *3*, 850-860
75. Moine, H.; Ehresmann, B.; Ehresmann, C.; Romby, P. *Probing RNA structure and function in solution*. In: *RNA structure and function*; Simons, R. W.; GrunbergManago, M. Eds. Cold Spring Harbor Laboratory Press: NY, 1997; pp 77-115
76. Sun, S.; Yoshida, A.; Piccirilli, J. A. *RNA* 1997, *3*, 1352-1363
77. Shpanchenko, O. V.; Dontsova, O. A.; Bogdanov, A. A.; Nierhaus, K. H. *RNA* 1998, *4*:1154-1164
78. Sontheimer, E. J.; Sun, S.; Piccirilli, J. A.. *Nature* 1997, *388*:801-805
79. Weinstein, L. B.; Jones, B. C.; Cosstick, R.; Cech, T. R. *Nature* 1997, *388*:805-808
80. Joseph, S.; Noller, H. F. *EMBO J* 1996, *15*:910-916
81. Cooper, M.; Johnston, L. H.; Beggs, J. D. *EMBO J* 1995, *14*:2066-2075
82. Heilek, G. M.; Noller, H. F. *RNA* 1996, *2*, 597-602
83. Beck, D. L.; Stump, W. T.; Hall, K. B. *RNA* 1998, *4*, 331-339
84. Hertzberg, R. P.; Dervan, P. B. *Biochemistry* 1984, *23*, 3934-3945
85. Wilson, K. S.; Noller, H. F. *Cell* 1998, *92*:131-139
86. Papavassiliou, A. G. *Biochem. J* 1995, *305*, 345-357
87. Darsillo, P.; Huber, P. W. *J Biol. Chem.* 1991, *266*, 21075-21082
88. Nagai, K.; Oubridge, C.; Jessen, T. H.; Li, J.; Evans, P. R. *Nature* 1990, *348*, 515-520
89. Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C. H.; Nagai, K. *Nature* 1994, *372*, 432-438
90. Nagai, K.; Oubridge, C.; Ito, N.; Avis, J.; Evans, P. *Trends Biochem. Sci.* 1995, *20*, 235-240
91. Varani, G.; Nagai, K. *Annu. Rev. Biophys. Biomol. Struct.* 1998, *27*:407-445
92. Nagai, K. *Curr. Opin. Struct. Biol.* 1996, *6*, 53-61
93. Bycroft, M.; Grünert, S.; Murzin, A. G.; Proctor, M.; St.Johnston, D. *EMBO J* 1995, *14*:3563-3571
94. Kharrat, A.; Macias, M. J.; Gibson, T. J.; Nilges, M.; Pastore, A. *EMBO J* 1995, *14*, 3572-3584
95. Varani, G. *Acc. Chem. Res.* 1997, *30*:189-195
96. Burd, C. G.; Dreyfuss, G. *Science* 1994, *265*:615-621
97. Siomi, H.; Matunis, M. J.; Micheal, W. M.; Dreyfuss, G. *Nucleic Acids Res.* 1993, *21*:1193-1198
98. Gibson, T. J.; Tompson, J. D.; Herringa, J. *FEBS letters* 1993, *324*:361-366
99. Siomi, H.; Choi, M.; Siomi, M. C.; Nussbaum, R. L.; Dreyfuss, G. *Cell* 1994, *77*:33-39
100. Hermann, H.; Fabrizio, P.; Raker, V. A.; Foulaki, K.; Hornig, H.; Brahms, H.; Luhrmann, R. *EMBO J* 1995, *14*:2076-2088

101. Seraphin, B. *EMBO J* **1995**, *14*:2089-2098
102. Kambach, C.; Walke, S.; Young, R.; Avis, J. M.; de la Fortelle, E.; Raker, V. A.; Luhrmann, R.; Li, J.; Nagai, K. *Cell* **1999**, *96*, 375-387
103. Lu, J.; Hall, K. B. *J Mol. Biol.* **1995**, *247*, 739-752
104. Sherly, D.; Boelens, W.; Van Venrooij, W. J.; Dathan, N. A.; Hamm, J.; Mattaj, I. W. *EMBO J* **1989**, *8*:4163-4170
105. Lutz Freyermuth, C.; Keene, J. D.; *Mol. Cell Biol.* **1989**, *9*, 2975-2982
106. Jessen, T. H.; Oubridge, C.; Teo, C. H.; Pritchard, C.; Nagai, K. *EMBO J* **1991**, *10*, 3447-3456
107. Bach, M.; Krol, A.; Luhrmann, R. *Nucleic Acids Res.* **1990**, *18*:449-457
108. Patton, J. R.; Habets, W. J.; Van Venrooij, W. J.; Pederson, T. *Mol. Cell Biol.* **1989**, *9*:3360-3368
109. Surowy, C. S.; van Santen, V. L.; Scheib-Wixted, S. M.; Spritz, R. A. *Mol. Cell Biol.* **1989**, *9*:4179-4186
110. Allain, F. H. T.; Howe, P. W. A.; Neuhaus, D.; Varani, G. *EMBO J* **1997**, *16*, 5764-5774
111. Hoffman, D. W.; Query, C. C.; Golden, B. L.; White, S. W.; Keene, J. D. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*:2495-2499
112. Howe, P. W.; Nagai, K.; Neuhaus, D.; Varani, G. *EMBO J* **1994**, *13*, 3873-3881
113. Hall, K. B. *Biochemistry* **1994**, *33*, 10076-10088
114. Hall, K. B.; Stump, W. T. *Nucleic Acids Res.* **1992**, *20*, 4283-4290
115. Stump, W. T.; Hall, K. B. *RNA* **1995**, *1*, 55-63
116. Nagai, K.; Oubridge, C.; Ito, N.; Jessen, T. H.; Avis, J.; Evans, P. *Nucleic Acids Symp. Ser.* **1995**, 1-2
117. Avis, J. M.; Allain, F. H.; Howe, P. W.; Varani, G.; Nagai, K.; Neuhaus, D. *J Mol. Biol.* **1996**, *257*, 398-411
118. Allain, F. H.; Gubser, C. C.; Howe, P. W.; Nagai, K.; Neuhaus, D.; Varani, G. *Nature* **1996**, *380*, 646-650
119. Kranz, J. K.; Lu, J.; Hall, K. B. *Protein Sci.* **1996**, *5*, 1567-1583
120. Boelens, W. C.; Jansen, E. J.; Van Venrooij, W. J.; Stripecke, R.; Mattaj, I. W.; Gunderson, S. I. *Cell* **1993**, *72*, 881-892
121. van Gelder, C. W.; Gunderson, S. I.; Jansen, E. J.; Boelens, W. C.; Polycarpou Schwarz, M.; Mattaj, I. W.; Van Venrooij, W. J. *EMBO J* **1993**, *12*, 5191-5200
122. Gunderson, S. I.; Beyer, K.; Martin, G.; Keller, W.; Boelens, W. C.; Mattaj, L. W. *Cell* **1994**, *76*, 531-541
123. Gunderson, S. I.; PolycarpouSchwarz, M.; Mattaj, I. W. *Mol. Cell* **1998**, *1*, 255-264
124. Gubser, C. C.; Varani, G. *Biochemistry* **1996**, *35*, 2253-2267
125. Jovine, L.; Oubridge, C.; Avis, J. M.; Nagai, K. *Structure* **1996**, *4*, 621-631
126. Pellizzoni, L.; Lotti, F.; Rutjes, S. A.; PierandreiAmaldi, P. *J. Mol. Biol.* **1998**, *281*, 593-608
127. Wolin, S. L.; Steitz, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81*, 1996-2000
128. Deutscher, S. L.; Harley, J. B.; Keene, J. D. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 9479-9483
129. Query, C. C.; Bentley, R. C.; Keene, J. D. *Mol. Cell Biol.* **1989**, *9*:4872-4881
130. Green, C. D.; Long, K. S.; Shi, H.; Wolin, S. L. *RNA* **1998**, *4*:750-765
131. Pruijn, G. J.; Slobbe, R. L.; Van Venrooij, W. J. *Nucleic Acids Res.* **1991**, *19*, 5173-5180
132. O'Brien, C. A.; Wolin, S. L. *Genes Dev.* **1994**, *8*, 2891-2903
133. Pruijn, G. J. M.; Simons, F. H. M.; vanVenrooij, W. J. *Eur. J. Cell Biol.* **1997**, *74*, 123-132
134. Wolin, S. L.; Steitz, J. A. *Cell* **1983**, *32*: 735.744
135. O'Brien, C. A.; Margelot, K.; Wolin, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*:7250-7254
136. van Gelder, C. W.; Thijssen, J. P.; Klaassen, E. C.; Sturchler, C.; Krol, A.; Van Venrooij, W. J.; Pruijn, G. J. *Nucleic Acids Res.* **1994**, *22*, 2498-2506
137. Farris, A. D.; O'Brien, C. A.; Harley, J. B. *Gene* **1995**, *154*, 193-198
138. Luhrmann, R.; Kastner, B.; Bach, M. *Biochim. Biophys. Acta* **1990**, *1087*:265-292
139. Lutz Freyermuth, C.; Query, C. C.; Keene, J. D. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*:6393-6397
140. Sherly, D.; Boelens, W.; Dathan, N. A.; Van Venrooij, W. J.; Mattaj, I. W. *Nature* **1990**, *345*:502-506
141. Bently, R. C.; Keene, J. D. *Mol. Cell Biol.* **1991**, *11*:1829-1839
142. Tsai, D. E.; Harper, D. S.; Keene, J. D. *Nucleic Acids Res.* **1991**, *19*:4931-4936
143. Nelissen, R. L.; Sillekens, P. T.; Beijer, R. P.; Geurts van Kessel, A. H.; Van Venrooij, W. J. *Gene* **1991**, *102*:189-196
144. Kwakman, J. H.; Konings, D. A.; Hogeweg, P.; Pel, H. J.; Grivell, L. A. *J Biomol. Struct. Dyn.* **1990**, *8*:413-430
145. Glickman, J. N.; Howe, J. G.; Steitz, J. A. *J Virol.* **1988**, *62*:902-911
146. Dock-Bregeon, A. C.; Westhof, E.; Giege, R.; Moras, D. *J Mol. Biol.* **1989**, *206*:707-722
147. Mougel, M.; Eyermann, F.; Westhof, E.; Romby, P.; Expert Bezancon, A.; Ebel, J. P.; Ehresmann, B.; Ehresmann, C. *J Mol. Biol.* **1987**, *198*:91-107
148. Celander, D. W.; Cech, T. R. *Biochemistry* **1990**, *29*:1355-1361
149. Williams, D. J.; Hall, K. B. *J Mol. Biol.* **1996**, *257*, 265-275
150. Varani, G. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*:379-404
151. Klein Gunnewiek, J. M. T.; Van de Putte, L. B.; Van Venrooij, W. J. *Clin. Exp. Rheum.* **1997**, *15*:549-560
152. Van Venrooij, W. J.; Hoet, R. M.; Castrop, J.; Hageman, B.; Mattaj, I. W.; Van de Putte, L. B. *J Clin. Invest.* **1990**, *86*:2154-2160
153. Hoet, R. M.; De Weerd, P.; Gunnewiek, J. K.; Koornneef, I.; Van Venrooij, W. J. *J Clin. Invest.* **1992**, *90*, 1753-1762
154. Tsai, D. E.; Keene, J. D. *J Immunol* **1993**, *150*, 1137-1145

155. StClair, E. W.; Burch, J. A. *Clin. Immun. Immunopath.*. 1996, *79*, 60-70
156. Hoet, R. M.; Kastner, B.; Luhrmann, R.; Van Venrooij, W. J. *Nucleic Acids Res.* 1993, *21*, 5130-5136
157. Billings, P. B.; Allen, R. W.; Jensen, F. C.; Hoch, S. O. *J Immunol* 1982, *128*:1176-1180
158. Habets, W. J.; Hoet, M. H.; De Jong, B. A.; Van der Kemp, A.; Van Venrooij, W. J. *J Immunol* 1989, *143*, 2560-2566
159. deWildt, R. M. T.; Finnern, R.; Ouwehand, W. H.; Griffiths, A. D.; vanVenrooij, W. J.; Hoet, R. M. A. *Eur. J. Immunol.* 1996, *26*, 629-639
160. Carmo Fonseca, M.; Tollervey, D.; Pepperkok, R.; Barabino, S. M.; Merdes, A.; Brunner, C.; Zamore, P. D.; Green, M. R.; Hurt, E.; Lamond, A. I. *EMBO J* 1991, *10*, 195-206
161. Hoet, R. M. A.; Raats, J. M. H.; de Wildt, R.; Dumortier, H.; Muller, S.; van den Hoogen, F.; Van Venrooij, W. J. *submitted* 1999,
162. Lindahl, L.; Zengel, J. M. RNase MRP and rRNA processing. *Mol. Biol. Rep.* 1996, *22*:69-73
163. Krol, A.; Westhof, E.; Bach, M.; Luhrmann, R.; Ebel, J. P.; Carbon, P. *Nucleic Acids Res.* 1990, *18*:3803-3811
164. Bohman, K.; Ferreira, J.; Santama, N.; Weis, K.; Lamond, A. I. *J Cell Sci.* 1995, *19*:107-113
165. Klein Gunnewiek, J. M. T.; Van Venrooij, W. J. *Autoantigens contained in the U1 small nuclear ribonucleoprotein complex*. In: *Manual of biological markers of disease*; van Venrooij, W. J.; Maini, R. N. Eds. Kluwer Academic Publishers: Dordrecht, 1994; pp 1-20
166. Carmo Fonseca, M.; Pepperkok, R.; Carvalho, M. T.; Lamond, A. I. *J Cell Biol.* 1992, *117*, 1-14
167. Matera, A. G.; Ward, D. C. *J Cell Biol.* 1993, *121*, 715-727
168. Teunissen, S. W. M.; vanGelder, C. W. G.; vanVenrooij, W. J. *Biochemistry* 1997, *36*, 1782-1789
169. Van Horn, D. J.; Eisenberg, D.; O'Brien, C. A.; Wolin, S. L. *RNA* 1995, *1*, 293-303
170. Pruijn, G. J.; Wingens, P. A.; Peters, S. L.; Thijssen, J. P.; Van Venrooij, W. J. *Biochim. Biophys. Acta* 1993, *1216*, 395-401
171. Itoh, Y.; Kriet, J. D.; Reichlin, M. *Arthritis Rheum.* 1990, *33*:1815-1821
172. O'Brien, C. A.; Harley, J. B. *EMBO J* 1990, *9*:3683-3689
173. Farris, A. D.; Koelsch, G.; Pruijn, G. J. M.; Van Venrooij, W. J.; Harley, J. B. *Nucleic Acids Res.* 1999, *27*: 4.1070-1078
174. Zuker, M. *Science* 1989, *244*:48-52
175. Rutjes, S. A. *submitted* 1999,
176. Pellizzoni, L.; Cardinali, B.; Lin Marq, N.; Mercanti, D.; Pierandrei Amaldi, P. *J Mol. Biol.* 1996, *259*, 904-915
177. Pellizzoni, L.; Lotti, F.; Maras, B.; Pierandrei Amaldi, P. *J Mol. Biol.* 1997, *267*, 264-275
178. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* 1994, *22*:4673-4680
179. Teunissen, S. W. M.; Stassen, M. H. W.; Pruijn, G. J. M.; Van Venrooij, W. J.; Hoet, R. M. A. *RNA* 1998, *4*: 9.1124-1133

Summary

Next to the well known messenger, ribosomal and transfer RNAs, a large number of small structural RNA molecules exist. These RNAs are bound to proteins, forming ribonucleoprotein particles (RNPs). RNPs are involved in many different processes in the cell, in particular pre-mRNA splicing and rRNA processing. For some of the RNPs the function is still unknown. An interesting feature of RNPs is that they are often targeted by autoantibodies occurring in patients suffering from an autoimmune disease. In order to better understand the function of these RNPs, it is important to study the structure of their RNA components. This thesis presents biochemical studies on the structures of three different RNA molecules which are in some way related to autoimmunity.

Chapter 1 introduces some basic principles of RNA structure and discusses some general biochemical methods that can be used to study the structure of RNA molecules. Additionally, the autoimmune related RNA and RNA-protein complexes studied in this thesis are discussed.

The U1A protein is part of the U1 snRNP complex, which plays an important role in the splicing of pre-mRNA and is an autoantigen in many patients suffering from systemic lupus erythematosus (SLE). The synthesis of U1A is auto-regulated: when U1A can no longer bind to the U1 snRNP (*i.e.* when there is an excess of U1A protein), it binds to the 3'UTR of its own pre-mRNA. As a consequence, the poly(A)-polymerase, which synthesises the poly(A)-tail that stabilises the pre-mRNA, is inhibited and the destabilised pre-mRNA will be degraded. Hence, the production of U1A protein will be reduced. Chapter 2 deals with the 3'UTR of U1A pre-mRNA and its complex with the U1A protein. The secondary structure of the naked RNA was studied using chemical structure probing techniques. The results confirmed and further refined the secondary structure proposed earlier. In addition, the complex of U1A with the pre-mRNA was studied with enzymatic probes. Apart from the protein binding sequence, the stem regions flanking these so-called Box sequences were found to be protected by the protein. Interestingly, a conformational change in the RNA molecule induced by protein binding was found.

Chapter 3 is also related to autoimmunity, but in a different way. It is known that patients that suffer from SLE sometimes produce autoantibodies directed to the RNA part of the U1 snRNP. It is difficult to study these RNA-autoantibody interactions because the sera of these patients are polyclonal. We therefore isolated a monoclonal anti-RNA antibody fragment from a phage display library, which was made from the IgG antibody repertoire of an autoimmune patient. This enabled us to study the complex of a single anti-RNA antibody and its target RNA, *i.e.* stem-loop II of U1 snRNA. The complex was studied using both enzymatic and chemical probes. The results showed that the centre of the RNA binding sequence of the antibody fragment consist of a number of non-canonical base pairs. These base pairs distort the regular helix and thus probably enable the antibody to bind in a base specific manner. Upon antibody binding the RNA showed a general stabilisation indicated by the stacking of three nucleotides in the loop structure of stem-loop II.

Chapter 4 describes a structure probing study on Y RNA molecules from several different organisms. Y RNAs are associated with Ro RNPs, which are autoantigens in Sjögren's syndrome and SLE. The cellular function of Ro RNPs is yet unknown. The human, frog, and iguana Y3 and Y4 RNAs were studied using chemical and enzymatic probes. The derived secondary structures are characterised by a conserved double-stranded region and a large flexible region in the central part of the RNA. The Ro60 and La protein, two Y RNA binding Ro RNP proteins, bind to the conserved double-stranded region, which constitutes the 'stable' part of the RNA. The flexible region was observed in all of the RNAs studied, indicating a conserved property rather than a conserved structure. These multiple conformations, either caused by dynamics or by multiple stable conformations (folding redundancy), are probably essential for the function of the RNP.

In Chapter 5 the results described in this thesis are discussed in a more general way. The finding that the unusual rather than canonical Watson-Crick and Wobble base pairing appears to influence RNA function is emphasised.

Samenvatting

Naast de wat bekendere messenger, ribosomale en transfer RNAs, bestaan er in cellen van zoogdieren zoals de mens een groot aantal kleine structurele RNAs. Deze RNA-moleculen worden gebonden door eiwitten en vormen zo *ribonucleoprotein particles* (RNPs). RNPs zijn betrokken bij een veelvoud aan functies binnen de cel, met name pre-mRNA splicing en rRNA processing. Van een aantal van de RNPs is er nog geen functie bekend. Een interessant aspect van deze kleine RNPs is dat sommige vaak het doelwit zijn van antilichamen die voorkomen in patiënten lijdend aan een auto-immuunziekte. Om de cellulaire functie van deze RNPs te kunnen begrijpen, is het belangrijk om de structuur die hun RNA-componenten hebben te bestuderen. Dit proefschrift beschrijft biochemisch onderzoek naar de structuur van drie verschillende RNA moleculen die allen in meer of mindere mate gerelateerd zijn aan auto-immuunziekten.

Hoofdstuk 1 presenteert een inleiding in de basisstructuren die een RNA molecuul kan bezitten en laat de verschillende biochemische methoden zien waarmee de structuur van RNA bepaald en bestudeerd kan worden. Verder worden de autoimmuun-gerelateerde RNA en RNA-eiwitcomplexen geïntroduceerd die in dit proefschrift zijn bestudeerd.

Het U1A eiwit maakt deel uit van het U1 snRNP complex, dat een belangrijke rol speelt bij de splicing van pre-mRNA en een autoantigeen is bij mensen die lijden aan systemische lupus erythematosus (SLE). The synthese van U1A is auto-gereguleerd: wanneer U1A niet meer in staat is om U1 snRNP te binden (als gevolg van een overmaat aan U1A), dan bindt U1A aan het 3'UTR gedeelte van het eigen pre-mRNA. Eenmaal gebonden voorkomt het U1A eiwit dat het poly(A)-polymerase, dat normaal een poly(A) staart synthetiseert en hiermee pre-mRNA stabiliseert, haar functie kan vervullen waardoor het pre-mRNA snel wordt afgebroken. Hierdoor zal de synthese van nieuw U1A eiwit worden verminderd. In hoofdstuk 2 wordt de 3'UTR van U1A pre-mRNA en haar complex met U1A bestudeerd. De secundaire structuur van het kale RNA is bepaald met behulp van specifieke chemische agentia. De resultaten hebben de eerder voorgestelde secundaire structuur bevestigd en verder verfijnd. Het complex met het U1A eiwit is bestudeerd met behulp van enzymen. Naast de eiwit-bindende sequenties (of Box sequenties), bleken grote delen van aangrenzende dubbelstrengs-gebieden door het eiwit beschermd te zijn. Bovendien werd er een verandering gevonden in de RNA-structuur, veroorzaakt door eiwit-binding.

Hoofdstuk 3 is op een andere manier gerelateerd aan autoimmuniteit. Het is bekend dat sommige SLE-patiënten autoantilichamen produceren tegen het RNA-gedeelte van U1 snRNP. Het is echter moeilijk om deze RNA-antilichaam-interacties te bestuderen, omdat sera van deze patiënten polyklonaal zijn en dus vele reactiviteiten vertonen. We hebben daarom een monoklonaal antilichaam-fragment geïsoleerd uit een faag-displaybank, die was gemaakt van het antilichaam-repertoire van een auto-immuunpatiënt. Op deze manier zijn we in staat geweest om een enkel anti-RNA antilichaam, gebonden aan haar RNA antigeen (stem-loop II van U1 snRNA), te bestuderen. Het complex is onderworpen aan enzymatisch en chemisch structuuronderzoek. De resultaten toonden aan dat het centrum van de

herkenningssequentie van het antilichaam-fragment bestaat uit niet-standaard basenparen. Deze ongebruikelijke basencombinatie veroorzaakt een vervorming van de normale dubbele helixstructuur en zorgt op die manier voor base-specifieke herkenning door het antilichaam-fragment. Het RNA laat een algemene stabilisatie zien als het antilichaam bindt, hetgeen resulteert in stapeling van drie nucleotiden in de loop-structuur van stem-loop II.

Hoofdstuk 4 beschrijft een onderzoek naar de structuur van Y RNA-moleculen van verschillende organismen. Y RNAs maken deel uit van Ro RNPs die een nog onbekende cellulaire functie hebben en autoantigeen zijn in Sjögren's syndroom en SLE. De Y3 en Y4 RNAs van mens, kikker en leguaan werden enzymatisch en chemisch behandeld om de structuur op te helderen. De secundaire structuren worden gekarakteriseerd door een geconserveerd dubbelstrengs gebied en een grote flexibele regio in het centrale gedeelte van het RNA. Het Ro60 en La eiwit, twee Y RNA-bindende Ro RNP eiwitten, associëren met het geconserveerde dubbelstrengs gebied, wat het 'stabiele' gedeelte van het RNA vormt. De flexibele regio werd gevonden in alle RNAs die bestudeerd zijn en kan beter een geconserveerde eigenschap worden genoemd in plaats van een geconserveerde structuur. Deze meervoudige conformaties, die of het gevolg zijn van de dynamiek van de RNA structuur of van verschillende stabiele structuren (*folding redundancy*), zijn waarschijnlijk essentieel voor de functie van het RNP.

In hoofdstuk 5 worden de resultaten die in dit proefschrift beschreven worden, in een groter kader bediscussieerd. Hierin wordt benadrukt dat juist de ongebruikelijke combinaties van basen in plaats van de canonische Watson-Crick en Wobble basenparen belangrijk zijn voor de functie van het RNA.

Dankwoord

Het dankwoord….. dat was het dan, de laatste (en misschien wel meest gelezen) letters die een bijzondere periode afsluiten. Bijzonder in veel verschillende opzichten, van "Hoezo is het experiment weer niet gelukt?" en "Ach, het is niet leuk als het meteen de eerste keer al lukt" tot het presenteren van experimenten die wel gelukt zijn op een groot congres. En van doorwerken tot in de late uurtjes tot doorwerken tot in de late uurtjes. Er zijn ook een aantal mensen die deze periode bijzonder hebben gemaakt (ook weer in veel opzichten) en sommigen zou ik daarvoor willen bedanken (niet allemaal: StafInfo bijvoorbeeld niet, terwijl zij het soms erg bijzonder maakten).

Walther, die het geheel heeft mogelijk gemaakt. Celia, die mij heeft geïntroduceerd in het mooie vak van 'proben' en Martijn voor zijn bijdragen aan hoofdstuk vier. De mensen bij het voormalig CAOS/CAMM centrum en in het bijzonder Hilbert voor je (soms goed beloonde) ondersteuning in de digitale biochemie.

Ger, bedankt voor je veeele goede suggesties, je tijd en je kritische blikken en vragen voor zowel het werk in het lab als voor het opschrijven van resultaten. Ik heb het altijd een plezier gevonden om met je samen te werken. Het was het wachten in de lange rij AiO's voor de deur wel waard!

De mensen van de VRT en diegenen die bij de VRT hebben gezeten en niet te vergeten Els, bedankt voor de gezellige sfeer zowel op het lab als daarbuiten. Dat is toch erg belangrijk voor het afronden van een promotie. Alle mensen met wie ik de kamer of het lab heb gedeeld: bedankt en excuses voor de soms wat luidruchtige (maar goede!!!) muziek die af en toe het lab in trilling bracht.

Arthur voor de computer ondersteuning en vooral de vele voorvallen ter lering ende vermaek. Ugh!

Gerard voor de prettige discussies over Jimmy versus Steve, Fender versus Ibanez en over hoe het eigenlijk allemaal zou moeten. Maar ook zeker de leuke en goede werkbesprekingen en niet te vergeten het jaarlijkse bezoek aan MP!

De rest van het lab wil ik bedanken voor de grote hoeveelheid aandacht die ik altijd weer kon verwachten ("De printer doet het niet!", "M'n muis is stuk", "Ik kan niet meer inloggen…", "Als ik hierop klik doet de mail het niet meer").

Virgil voor zijn nooit aflatende belangstelling en het altijd moeten aanhoren van: "Ja, druk he? En met jou?". Misschien dat ik hierna de UV777 eens in actie kan laten komen!

Ik vergeet altijd wel iets, dus voor diegenen een niet minder gemeend: Thanx!

Aviva, jij bent degene aan wie ik het meeste verschuldigd ben. Bedankt voor al je steun, zonder jou was dit hele boekje er waarschijnlijk niet eens gekomen!

Curriculum Vitae

Sander Teunissen werd op 21 september 1971 geboren te Nijmegen. In 1989 behaalde hij het VWO diploma aan het Dukenburg College, waarna de studie scheikunde werd gevolgd aan de Universiteit van Nijmegen. Tijdens deze studie heeft hij twee hoofdrichtingen afgerond: organische chemie bij het CAOS/CAMM centrum onder leiding van dr. J. W. Boiten/ dr. M. A. Ott en biochemie bij Prof. dr. W. J. van Venrooij, onder leiding van dr. C. W. G. van Gelder. Het doctoraal examen werd behaald in 1994.

Vanaf eind 1994 is hij AiO geweest bij de afdeling Biochemie van prof. dr. W. J. van Venrooij onder leiding van dr. G. J. M. Pruijn. Hij heeft onderzoek verricht naar de structuur van verschillende auto-immuun gerelateerde RNA moleculen, beschreven in dit proefschrift. Vanaf januari 1999 is hij werkzaam bij ORIGIN in Nijmegen.